An Alternative Approach to Visualizing Stock Market Correlation Matrices

- An Empirical study of forming portfolios that contain only small numbers of stocks using both existing and newly discovered visualization methods

A Thesis Submitted to the Department of Economics and Finance

University of Canterbury

By

Cheng Juan Zhan

In Partial Fulfillment of the Requirements of
the Degree of Master of Commerce in Finance

July 2014

TABLE OF CONTENTS

Declaration

This thesis is a presentation of my original work. Wherever contributions of others are involved, every effort is made to indicate this clearly, with due reference to the literature, and acknowledgement of collaboration and discussions.

The work was done under the guidance of Doctor Bill Rea of the University of Canterbury, Christchurch, New Zealand and Doctor Alethea Rea of Data Analysis Australia, Perth, Australia.

(Hannah) Cheng Juan Zhan

Department of Economics and Finance

University of Canterbury

11th July 2014

Acknowledgements

I am indebted and thankful to Doctor Bill Rea and Doctor Alethea Rea for their excellent supervision, guidance, programming support, patience and encouragement during the ups and downs of this research.

I would like to thank Professor Glenn Boyle and the department of Economics and Finance for their hardware support. I would also like to thank Economics and Finance library liaison Cuiying Mu for her help with data collection and efficient use of the EndNote. I would also like to thank the examiners for spending time to examine this thesis.

Abstract

The core of stock portfolio diversification is to pick stocks from different correlation clusters when forming portfolios. The result is that the chosen stocks will be only weakly correlated with each other. However, since correlation matrices are high dimensional, it is close to impossible to determine correlation clusters by simply looking at a correlation matrix. It is therefore common to regard industry groups as correlation clusters. In this thesis, we used three visualization methods namely Hierarchical Cluster Trees, Minimum Spanning Trees and neighbor-Net splits graphs to "collapse" correlation matrices' high dimensional structures onto two-dimensional planes, and then assign stocks into different clusters to create the correlation clusters. We then simulated sets of portfolios where each set contains 1000 portfolios, and stocks in each of the portfolio were picked from the correlation clusters suggested by each of the three visualization methods and industry groups (another way of determine correlation clusters). The mean and variance distribution of each set of 1000 simulated portfolios gives us an indication of how well those clusters were determined.

The examinations were conducted on two sets of financial data. The first one is the 30 stocks in the Dow Jones Industrial average which contains relatively small number of stocks and the second one is the ASX 200 which contains relatively larger number of stocks. We found none of the methods studied consistently defined correlation clusters more efficiently than others in out-of-sample testing.

The thesis does contribute the finance literature in two ways. Firstly, it introduces the neighbor-Net method as an alternative way to visualize financial data's underlying structures. Secondly, it used a novel "visualization" approach to portfolio diversification.

Key words: neighbor-Net, cluster analysis, splitsTree4, Minimum spanning tree, Hierarchical cluster.

Chapter 1 Introduction and motivation

1.1 The power of visualization

The saying "a picture is worth a thousand words" refers to the notion that complex information or ideas can be understood by conveying them in a picture. It seems, to the author, still a mystery how complex information or data can be understood easily in image form. At first glance, one may argue that languages are too limited to express the precise meanings of complex ideas or that human brains are overwhelmed when trying to comprehend large inflows of information and cannot grasp the structure of a massive data set at a glance until they can see it in a visual form. Either way, the saying effectively expresses the enormous power of visualization.

Visualisation is an essential tool that has allowed humans to understand phenomena from the dawn of human history. Long before the known existence of written languages, humans were using pictures to communicate with each other. In fact, visual objects such as paintings and sculptures were the main tools used to understand aspects of ancient civilizations when our knowledge of ancient languages is still very limited. Figure 1.1 shows a picture of the box named "the standard of Ur" which was discovered in the early twentieth century in south western Iraq. Even though the use of the box is still unknown, the frescos on both sides of the box reveal some crucial aspects of life in this ancient city, including information on its social hierarchy, agriculture, technology, occupations and religious conduct.

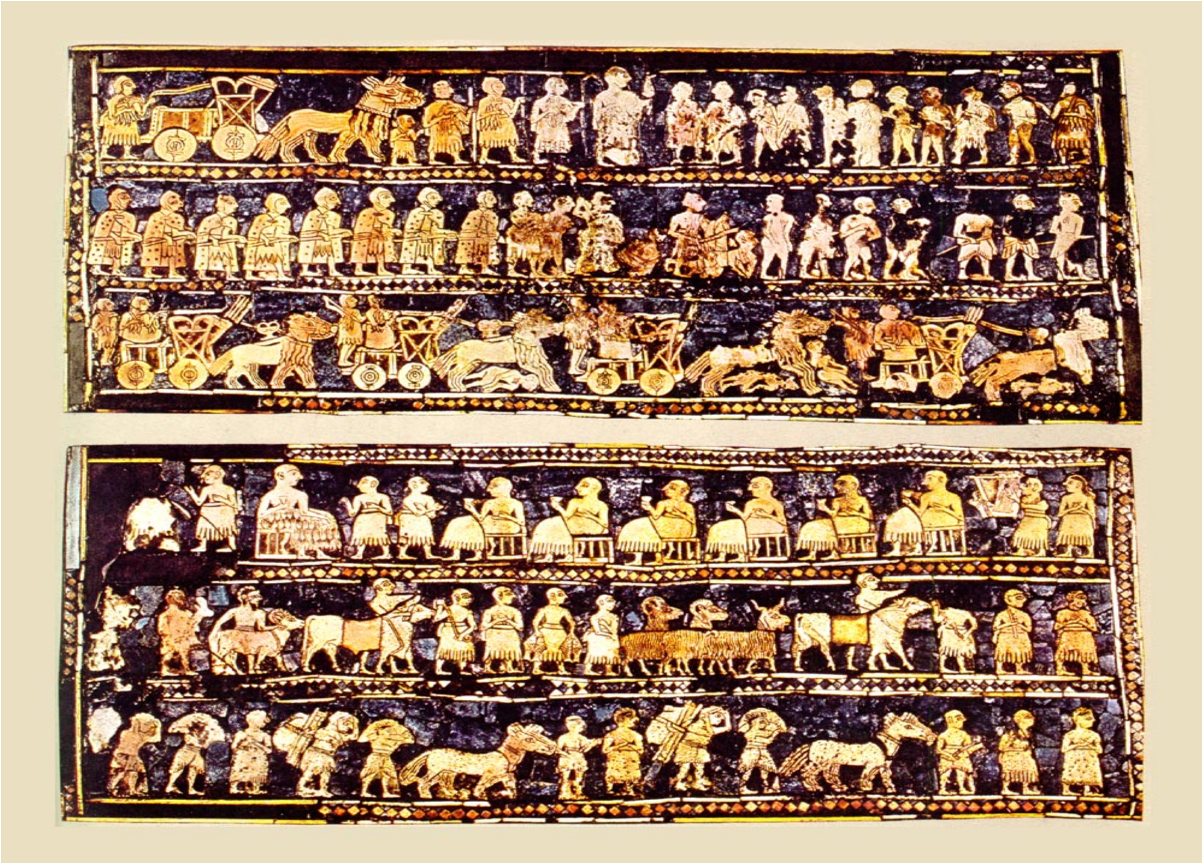


Figure 1.1 The standard of Ur.

1.2 Structure discovery through the data visualization approach

Since the 1980s, with the help of super computers, the world has experienced first-hand the power of being able to simulate physical systems. These simulations generate large amounts of data as their output. Data visualisation, which produces visual aids using large data sets, has gradually extended into almost every area of business and research. If data can be successfully gathered and expressed as graphs or images without much loss of information, the realisation of data's properties and structures can become almost immediately apparent. A variety of software packages have been developed in the past decades to improve the sophistication of the visualisation approaches applied in different fields. Examples of these software packages include DBMiner "*for interactive mining of multiple-level knowledge in large relational databases*" (http://www.dbminer.com); Spotfire "*which provides dynamic* *statistical analysis, particularly for businesses that offer enterprise-strength self-service predictive analytics to speed up decision-making and help customers achieve a two-second advantage™* " (Retrieved from http://spotfire.tibco.com) and WinViz for visualization of multidimensional data (Ong et al. 1996).

Visualisation software has been successfully applied in many fields. An example of its effective use is seen in the field of crime investigation. Police forces use geographic information systems (GIS) to analyse spatial patterns such as the locations and frequencies of the crimes. Points may be plotted on a map to detect patterns, clusters or randomness of the crimes. Figure 1.2 is an example of non-serious crime analysis as a 3-dimensional graphic. The clusters and structures of the crime are easily observed from the graph and such information is crucial for police resource allocation and forecasting.

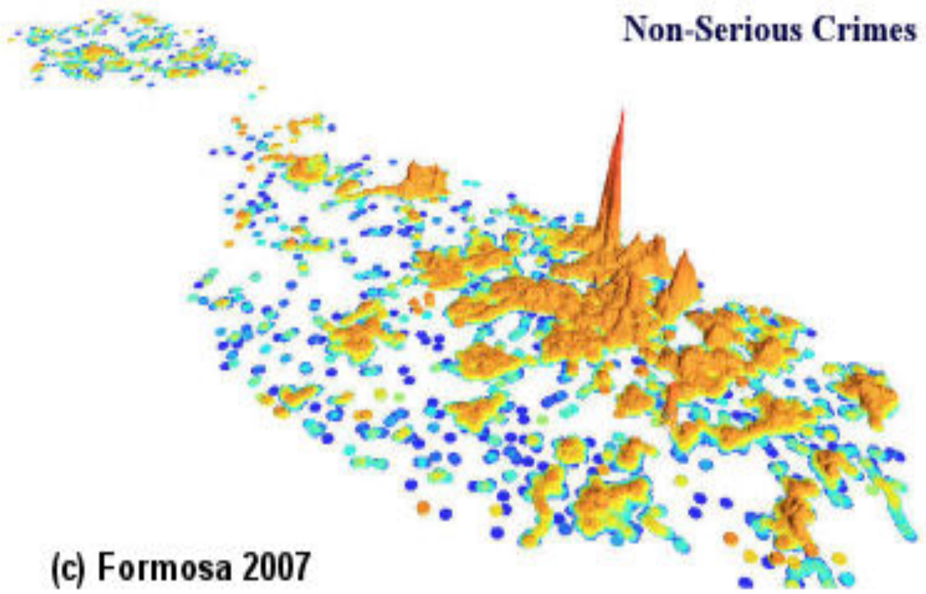


Figure 1.2 A 3-D analysis of non-serious crimes retrieved from http://www.crimemalta.com/map.htm.

Another example of a great use of the visualisation approach is the stream graphic used in social network analysis. Figure 1.3 shows the key words associated with the search word "earthquake" as at 11:40, 16th April 2013. While it is understandable to a New Zealander that words such as Japan and region are closely associated with the search word "earthquake", it is surprising to see the words Boston and Iraq also appearing with high frequency. These kinds of phenomena, when expressed in raw data form, are very hard for humans to comprehend the structures of the information hidden in the data set. Yet, the stream graph reveals immediately the structure of the data and is able to give the readers a clearer picture and therefore better understanding. Such a stream graph can be utilised in many fields. For instance, a shampoo manufacturer, prior to creating new products, can use the search to find out what words are related to the word shampoo, and then analyse the correlations behind their appearance to adjust their production, packaging and/or marketing strategies. For a simple example, the word "eye" appears together with the word shampoo with high frequency, and after looking into the search sentences that contain the word "eye", most of them were about shampoo going into the eyes and causing discomfort. Therefore, the manufacturer may consider launching a product that is less irritating to the eyes.

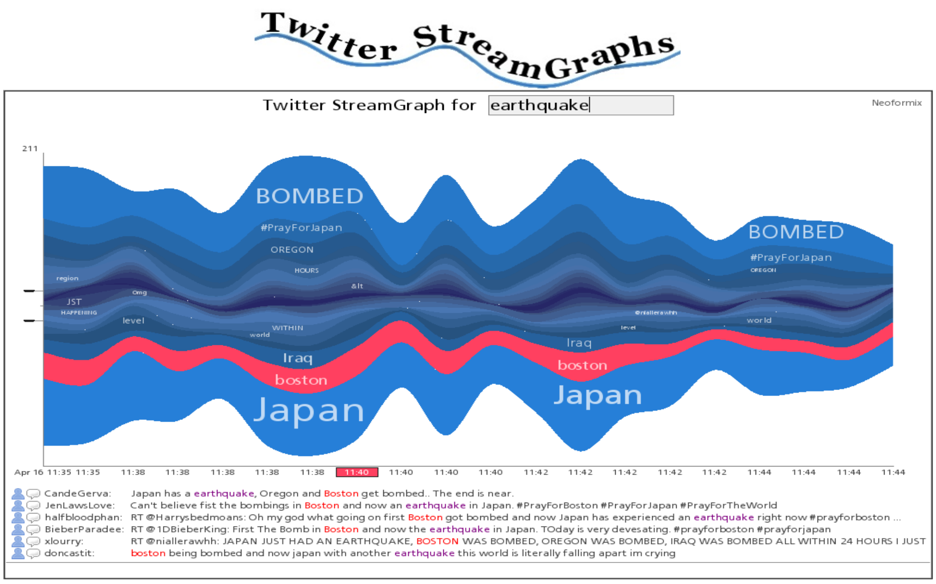


Figure 1.3. A Stream graph of latest 1000 tweets which contain the search word "earthquake" at 11:40 16th April 2011.

1.3 The existing visualization methods in the finance literature

The idea of *finance* has been in existence since the dawn of human civilization which is borrowing and lending at an interest. Prior to 1950s, graphical financial data analysis was largely conducted within a 2-dimensional plane. For instance, a line graph shows different types of mortgage rates across time can be used for observing the trends of their movements over time (Figure 1.4). Since there are only two variables, the mortgage rate and time, the graph adequately demonstrates the movements of the rate if the researcher wishes to observe the trend.

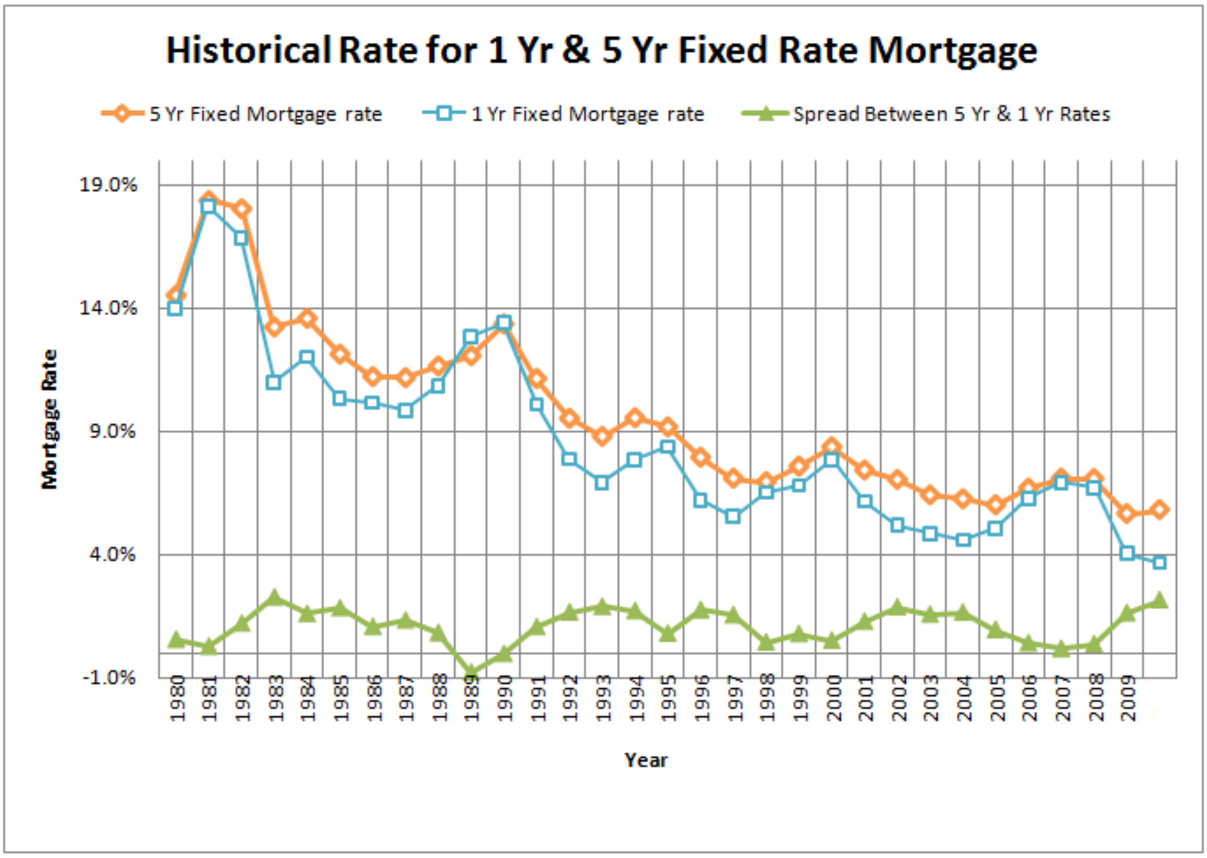


Figure 1.4. Historical rate for 1 year and 5 year fixed rate mortgages in Canada for the period 1980 to 2009.

Another example of 2-dimensional data visualization technique is the scatter plot matrix in Figure 1.5. A scatter plot matrix presents the degree of correlation between any two variables within a set of n variables. Figure 1.5 contains a simple scatter plot matrix that suggests the correlations among 3 variables, x, y and z. For example, the sub-graph on row 2 column 3 indicates the correlation between the variable y and the variable z, with the variable y on the vertical axis and the variable z on the horizontal axis. There is a positive correlation between variable y and variable z as suggested by the sub-graph. The scatter plot matrix effectively ensures that each sub-graph shares the same scale. That is, along each column or row of the matrix, one variable is kept the same while other variables are changed in each successive plot. This may help users to detect any pattern within the data sets when looking along the columns or rows. However, Everitt (Everitt, 1978, p. 5) stated: "*that scatter grams for all pairs of variables might be examined as a simple method of "looking" at the data, but although this approach may be useful in some situations it is, in general, very unsatisfactory for two reasons. Firstly, if the number of variables is greater than about ten then the number of such plots to be examined is large, and such examination is as likely to lead to confusion as enlightenment about the structure of the data. Secondly, such plots may be very misleading since any structure present in the original p-dimensional space of the data is not necessarily reflected by that present in scatter grams of pairs of variables*". Despite the limitations of the scatter plot approach, it is still the main method when it comes to the representation of correlations in the finance literature (Redpath, 2000).

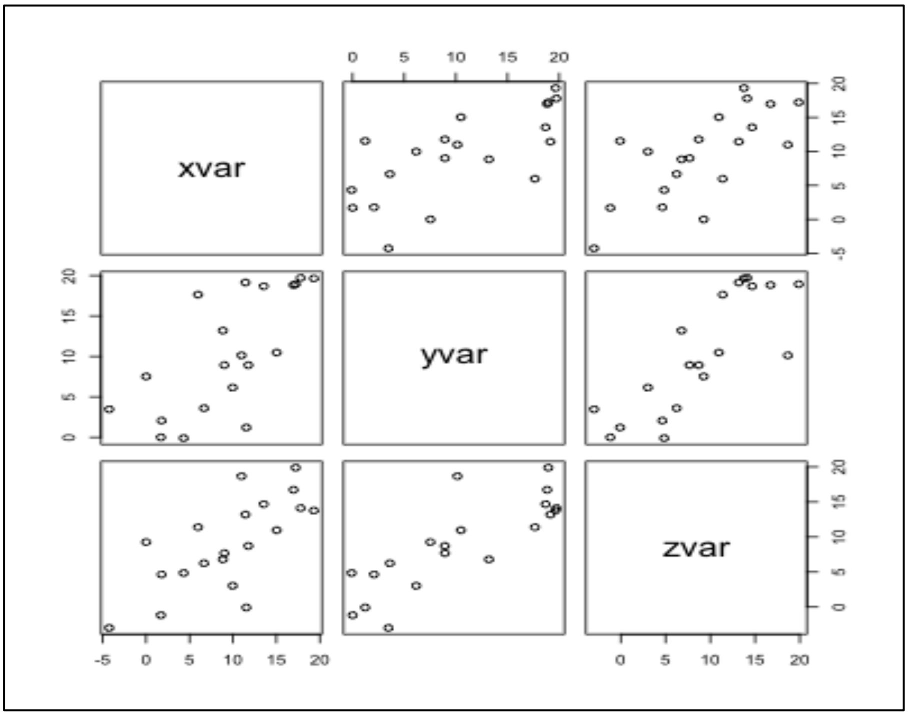


Table 1.5. An example of scatter plot matrix.

1.4 The objectives and structure of this thesis

The objective of this thesis is to utilize the power of visualization in financial data analysis. In particular, we used three visualization methods namely; the Hierarchical Cluster Tree (HCT), the Minimum Spanning Tree (MST) and the neighbor-Net splits graph to define clusters that lie within correlation matrices. We then simulated sets of portfolios; each portfolio contains two, four or eight stocks which were picked from different clusters suggested by these three different methods. I hoped to find a better visualization method that reveals the true structure of the correlation matrices more efficiently than other methods.

The structure of the rest of this thesis is as follow: Chapter 2 reviews the literature. The literature was split into two parts. Firstly, we briefly introduce the two visualization methods namely the HCT and the MST that exist in the finance literature and use small 6 by 6 matrices to demonstrate how the HCT and the MST were constructed. Secondly, we review the current problem faced by investors that hold only a small number of stocks in their portfolios. The two parts together form our incentives of why we conducted this particular research. Chapter 3 introduces the “new” visualization method – the neighbor-Net method. Chapter 4 describes four ways of transforming correlation matrices into distance matrices. In Chapter 5, we preliminarily examine the three visualization methods using a relatively small data set-Dow Jones 30 Industrial index and in Chapter 6 we examine only the neighbor-Net method on a relatively larger dataset. Chapter 7 contains some discussion about using neighbor-net to analyse systematic and idiosyncratic risk. Chapter 8 contains the conclusions and suggestions for further research.

Chapter 2 Literature review on multivariate data visualizations and modern portfolio diversifications

One of the purposes of this thesis is to visualize and detect the structures, if any, in large stock correlation matrices and then "group" the stocks into different clusters to evaluate the performance of portfolios formed by selecting stocks based on these correlation clusters. While the detailed research questions will be given in Section 2.3, it is appropriate to firstly review the two areas that form the basis of our research questions. Hence, in Section 2.1, we will review the visualization of clusters using multivariate data. Within this section, we first illustrate the challenges of visualizing multivariate data in Section 2.1.1. Then, because the cluster analysis is a subset of the field of multidimensional scaling, we will review methods of multidimensional scaling in Section 2.1.2 and the field of cluster analysis will be reviewed in Section 2.1.3. In Section 2.2, we review the mean-variance portfolio selection theory and other portfolio selection theories based on risk allocation and their limitations and applications. Finally in Section 2.3, the research questions emerge based on Section 2.1 and Section 2.2.

2.1 Cluster analysis of multivariate data using visualization

2.1.1 The challenge of visualizing multivariate data

It is important for researchers to be able to understand what, if any, structures lay within a correlation matrix and more importantly the change of the correlation structures over time. However, correlations of multivariate data are multidimensional and it is not feasible to express its structure on a two-dimensional plane. We use a five-variable correlation matrix as an example to explain the difficulty in direct visualization of high dimensional data sets. Table 2.1 contains a five-variable correlation matrix which was transformed into distance matrix using the formula 1-correlation (Table 2.2). The distances AB, AC and BC can be easily drawn on the two-dimensional plane XY as shown in Figure 2.1. However, there is no point D on the plane XY whose distance to A, B and C are exactly 1.37, 1.32 and 1.91 respectively. To find the correct position for point D, one has to "elevate" point D to a third dimension – the Z axis. It is only in this third dimension that we can find a position for D whose distance to A, B and C are exactly 1.37, 1.32 and 1.91. Furthermore, a correct position for point E may or may not exist in this three-dimensional space, if point E cannot be actually represented in this space, then a fourth dimension must be added to show point E correctly.

Stocks	A	B	C	D	E
A	—				
B	.87	—			
C	.64	.56	—		
D	.37	.32	.91	—	
E	.93	-.35	.54	-.43	—

Table 2.1. An example of a correlation matrix.

Stocks	A	B	C	D	E
A	—				
B	.13	—			
C	.36	.44	—		
D	.63	.68	.09	—	
E	.07	1.35	.46	1.43	—

Table 2.2. The new matrix after each correlation is deducted from 1. This matrix represents dissimilarities, that is, a larger number represents a larger dissimilarity.

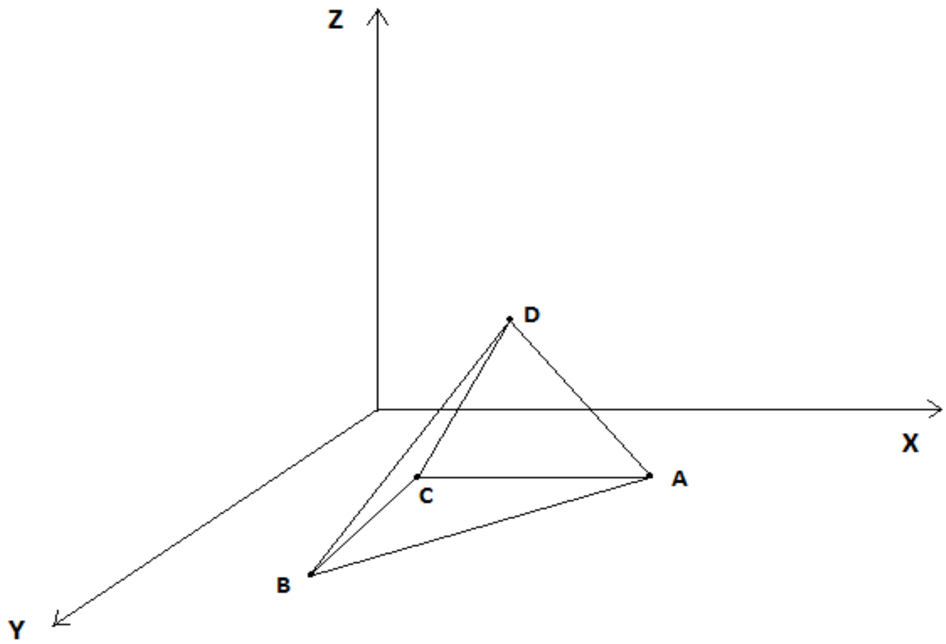


Figure 2.1. Points A, B, C and D in a three-dimensional space. It may not be feasible to represent point E in table 2.2 in this space.

Therefore, as we can see, the structure of even a small correlation matrix is very hard to visualize. When the number of stocks in a correlation matrix increases to, for example 30, then the number of correlations to be estimated is $\frac{n\times(n-1)}{2} = \frac{30\times 29}{2} = 435$. This requires up to (n-1) - dimensional space to represent all the distances. In the 30 stocks case, up to 29 dimensions are required. Hence, it becomes extremely hard for human eyes to observe its structure at a glance.

2.1.2 The fundamentals of multidimensional scaling and cluster analysis

2.1.2.1 Metric Multidimensional scaling

To be able to visualize the n-dimensional structure among the variables, the broad field of multidimensional scaling (MDS) techniques are used.

"Multidimensional scaling is a method that represents measurements of similarities (or dissimilarities) among pairs of objects as distances between points of a low- dimensional multidimensional space" - Borg and Groenen (2005). In other words, the multidimensional scaling procedure is trying to show multidimensional distances on a lower two or three-dimensional space.

It is worth noting that multidimensional scaling covers a wide range of techniques with different purposes. There are two main types of multidimensional scaling algorithms (Florian 2003). The first type is metric multidimensional scaling where the inputs are matrices of known distances. The second type is non-metric multidimensional scaling where there are non-parametric monotonic relationships between the dissimilarities. That is, the distance in the matrices cannot be measured numerically. Since this thesis focuses particularly on finding the clusters or groups of items with similarities or dissimilarities, we briefly review the classic metric MDS and its applications.

Principle component analysis also called classical MDS

A number of researchers have tried to use MDS in financial data analysis. For example, Rankin (2003) tried to use principle component analysis to improve portfolio selection; Driessen et al. (2003) used factor analysis to analyse international bond returns and Deboech and Kohonen (1998) used self-organising maps to reduce the dimensions of financial data to better understand the financial markets.

2.1.2.2 Cluster analysis

Once a high dimensional data set can be visualized, the next step for the purpose of our research is to “group” the stocks into different clusters. A crucial part of this thesis is to “collapse” the high dimensional correlation matrices to a lower dimension and then group stocks into different clusters. So it is only appropriate that after giving an overview of MDS, we now briefly review two cluster analysis methods which are used in financial data analysis.

The hierarchical clustering tree (HCT)

The hierarchical clustering analysis is one of the most commonly used procedures in cluster analysis. The idea behind it is to build a binary tree of the data which successively groups points that are similar to each other. There are different types of algorithms which can be used in creating a hierarchical tree, and we will only concentrate the agglomerative algorithm for the purpose of this thesis. The agglomerative clustering method (bottom up method) starts by treating each data point as a cluster that contains a single point, the so-called singleton or atomic cluster, then merges these atomic clusters into larger and larger clusters.

There are a number of different algorithms for merging clusters such as single linkage algorithm, complete linkage algorithm and average linkage algorithm (Kaufman et al., 1990). For the purpose of this paper, we are going to illustrate two of these algorithms namely single linkage algorithm and average linkage algorithm with examples.

Single linkage clustering:

A single linkage clustering method forms clusters by searching for the smallest distance between any point in one cluster and any one point in another cluster starting from atomic clusters. That is, the distance of the two clusters to be merged is

$$d_{G,H} = \min_{\substack{i \in G \\ j \in H}} d_{ij}$$

where d is the distance between clusters G and H, points $i \; and \; j$ that satisfy the above equation. This can be illustrated in Figure 2.2.

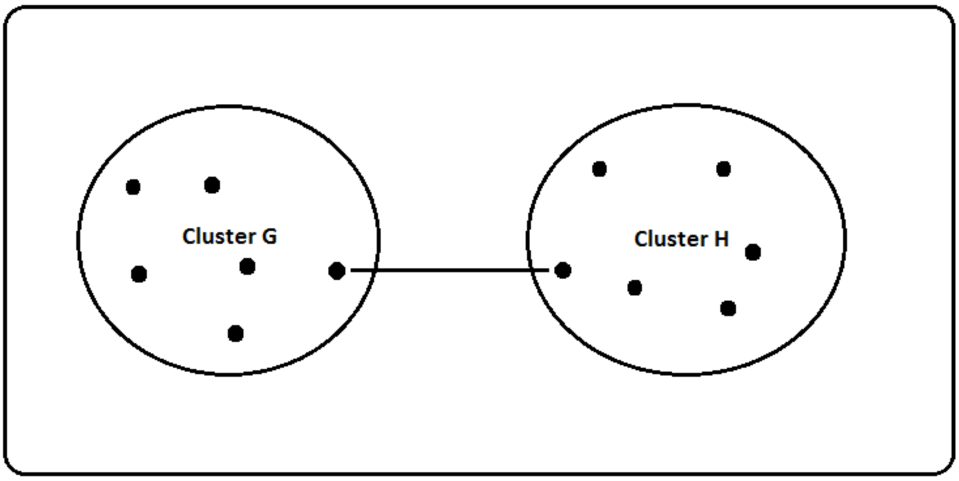


Figure 2.2. Single linkage clustering which combines two clusters as long as the distance between any one point in one cluster for example cluster G and any one point in the other cluster, for example cluster H, is the shortest. Therefore, cluster G and cluster H are combined.

Group Average linkage method:

A group average linkage method merges the two clusters that have the closest distance on average. The distance between clusters is given by $d_{G,H} = \frac{1}{N_G N_H} \sum_{i \in G} \sum_{j \in H} d_{ij}$ where cluster G and Cluster H are the two clusters to be merged. This can be illustrated as in Figure 2.3.

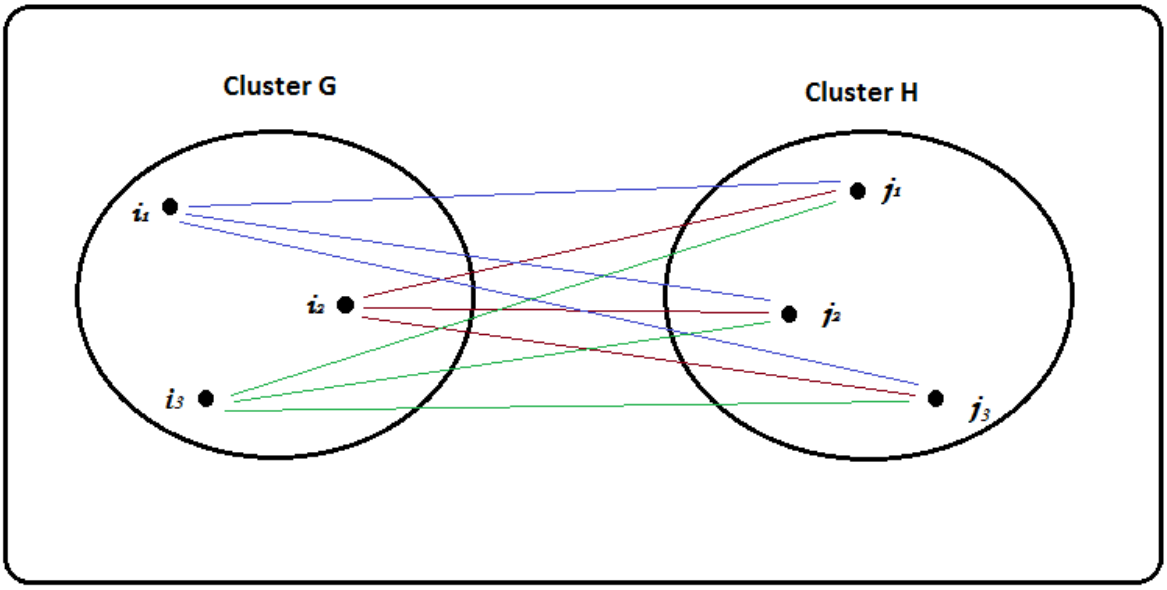


Figure 2.3. Group average linkage method merges the two clusters that have the closest distance on average.

Between the single linkage clustering and the group average clustering, the group clustering is the better method of finding groups with similarity, because group average clustering takes into account the distance between all points in two cluster groups that are to be merged.

A demonstration of the group averaging clustering method:

Table 2.3 contains correlation metric that shows the correlation among six stocks from Dow Jones 30 industrial index. We then transform the distances among all stocks using the formula **1-Correlation** shown in Table 2.4. By using the group average algorithm, we first find the two stocks that have the smallest distance, each stock is regarded a singleton cluster at this stage. It is easy to see that JPM and BOA are the two closest two stocks, therefore we merge the first two singleton clusters JPM and BOA (Refer to Figures 2.4 and 2.5). To find the next closest two groups, we need to calculate the average distance between points in each cluster group.

	AA	XOM	AXP	BOA	IBM	JPM
AA	1	0.400	0.514	0.403	0.537	0.288
XOM	0.400	1	0.308	0.249	0.427	0.239
AXP	0.514	0.308	1	0.569	0.549	0.573
BOA	0.403	0.249	0.569	1	0.440	0.767
IBM	0.537	0.427	0.549	0.440	1	0.551
JPM	0.288	0.239	0.573	0.767	0.551	1

Table 2.3. Correlation matrix of six stocks in the Dow Jones Industrial Average.

	AA	XOM	AXP	BOA	IBM	JPM
AA	0	0.600	0.486	0.596	0.463	0.712
XOM	0.600	0	0.692	0.751	0.573	0.761
AXP	0.486	0.692	0	0.431	0.451	0.427
BOA	0.596	0.751	0.431	0	0.560	**0.233**
IBM	0.463	0.573	0.451	0.560	0	0.450
JPM	0.712	0.761	0.427	**0.233**	0.450	0

Table 2.4. The correlation matrix transformed into a distance based matrix using the formula 1-correlation. The number in bold is the distance between the closest two stocks.

$$d_{BOA/JPM-AA} = \frac{1}{2}\left(d_{BOA,AA} + d_{JPM,AA}\right) = \frac{1}{2}\,(0.596 + 0.712) = 0.654$$

$$d_{BOA,JPM-XOM} = \frac{1}{2}\left(d_{BOA-XOM} + d_{JPM,XOM}\right) = \frac{1}{2}\,(0.751 + 0.761) = 0.756$$

$$\boldsymbol{d_{BOA,JPM-AXP=}\frac{1}{2}\left(d_{BOA-AXP} + d_{JPM-AXP}\right) = \frac{1}{2}\,(0.431 + 0.427) = 0.429}$$

$$d_{BOA,JPM-IBM} = \frac{1}{2}\left(d_{BOA,IBM} + d_{JPM,IBM}\right) = \frac{1}{2}\,(0.560 + 0.450) = 0.505$$

	AA	XOM	AXP	BOA/JPM	IBM
AA	0	0.600	0.486	0.654	0.463
XOM	0.600	0	0.692	0.756	0.573
AXP	0.486	0.692	0	**0.429**	0.451
BOA/JPM	0.654	0.756	**0.429**	0	0.505
IBM	0.463	0.573	0.451	0.505	0

Table 2.5. The new distances among all stocks after combining BOA and JPM. The closest two stocks are BOA/JPM and AXP with a distance of 0.429.

The clusters that are closest to each other from above distances are cluster BOA-JPM and cluster AXP (Table 2.5). Therefore JPM-BOA and AXP form the next cluster. The average distance between points in cluster JPM-BOA-AXP and points in other clusters are:

$$d_{BOA/JPM/AXP-AA} = \frac{1}{3}\left(d_{BOA,AA} + d_{JPM,AA} + d_{AXP,AA}\right) = \frac{1}{3}(0.596 + 0.712 + 0.486) = 0.598$$

$$d_{BOA/JPM/AXP-XOM} = \frac{1}{3}(d_{BOA,XOM} + d_{JPM,XOM} + d_{AXP,XOM}) = \frac{1}{3}(0.751 + 0.761 + 0.692)$$

$$= 0.7347$$

$$d_{BOA/JPM/AXP-IBM} = \frac{1}{3}\left(d_{BOA,IBM} + d_{JPM,IBM} + d_{AXP,IMB}\right) = \frac{1}{3}(0.560 + 0.450 + 0.451)$$

$$=0.487$$

	AA	XOM	BOA/JPM/AXP	IBM
AA	0	0.600	0.598	**0.463**
XOM	0.600	0	0.7347	0.573
BOA/JPM/AXP	0.598	0.7347	0	0.487
IBM	**0.463**	0.573	0.487	0

Table 2.6. The new distances among all stocks after combining BOA, JPM and AXP. The closest two stocks are AA and IBM with a distance of 0.463.

Taking into account all the distances, the distance between AA and IBM are the next closest clusters (Table 2.6). The algorithm continues until all there is finally one big cluster that contains all the points in the data set. The final result can be expressed in either as a Venn diagram as in Figure 2.4 or as a dendrogram in Figure 2.5. For the simplicity, we will use dendrogram to visualize the hierarchical clustering in the remainder of this thesis.

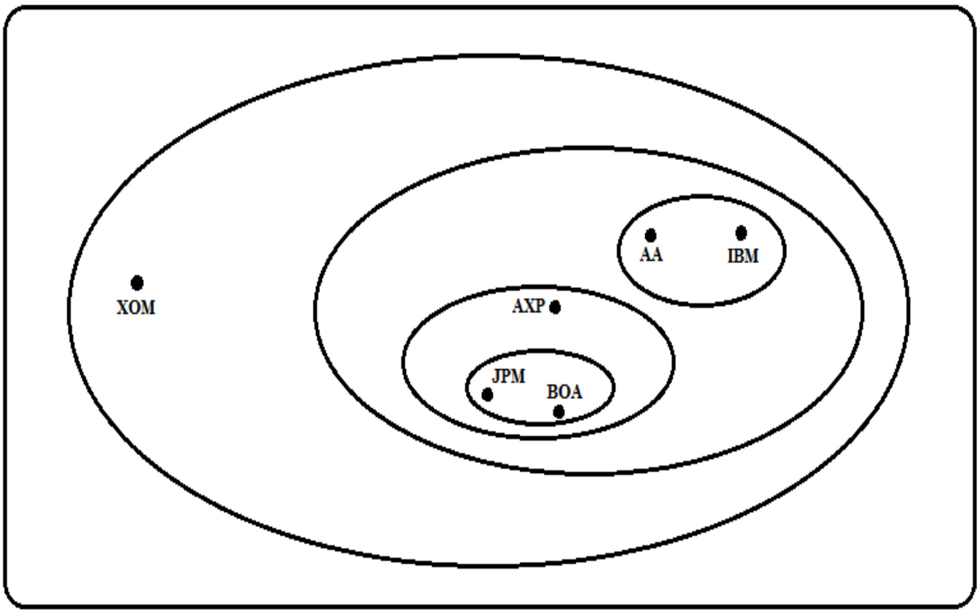


Figure 2.4. Venn diagram indicating clustering of the six stocks.

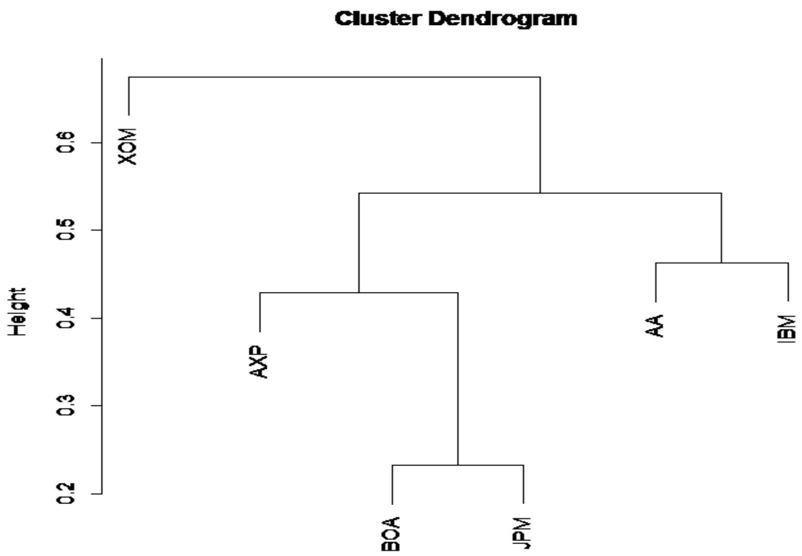


Figure 2.5. Tree-like dendrogram indicate the clusters of the six stocks.

The drawback of the hierarchical clustering is that it “forces” a tree-like structure of the data even when that structure is not tree-like.

The minimum spanning tree (MST)

Another technique is called a non-linear mapping or minimum spanning tree. The technique “seeks to obtain a low dimensional representation of p-dimensional multivariate data by seeking a p*-dimensional configuration which minimises the expression E given by

$$E = \frac{1}{\sum_{i<j} d_{ij}} \sum_{i,j} \frac{(d_{ij} - d^{*}{}_{ij})^2}{d_{iij}}$$

where d_{ij} is the Euclidean distance between observations i and j in the original p-dimensional space, and d_{ij}^{*} is the Euclidean distance between the p*-dimensional points representing these observations in the lower dimensional space” (Kruskal and Wish 1978).

As shown in Tables 2.7 through 2.11 and Figures 2.7 through 2.11, the six by six distance based correlation matrix builds MST as follows:

Since it does not matter which stock to start with, we will start building the MST from stock AA. The distances between AA and all other stocks are compared; we found that the shortest distance is 0.463 (in red) between AA and IBM. Therefore, AA and IBM are connected in Figure 2.7. Since AA and IBM are in the map, the next step is to consider the distances between AA and all other stocks as well as the distances between IBM and all other stocks. As shown in Table 2.8, the distances in bold and black are the ones to be considered. As shown in Table 2.8 and Figure 2.8, the closest distance is 0.450 between IBM and JPM therefore is connected. As shown in Tables 2.9 through 2.11and Figures 2.9 through 2.11, the MST continued to evolve until all stocks were connected.

	AA	XOM	AXP	BOA	IBM	JPM
AA	**0.00**	**0.600**	**0.486**	**0.596**	**0.463**	**0.712**
XOM		0.00	0.692	0.751	0.573	0.761
AXP			0.00	0.431	0.451	0.427
BOA				0.00	0.560	0.233
IBM					0.00	0.450
JPM						0.00

Table 2.7. For convenience, the MST is started from AA. The distances between AA and all other stocks are compared; the stock that is closest to AA is IBM with a distance of 0.463 (in bold red).

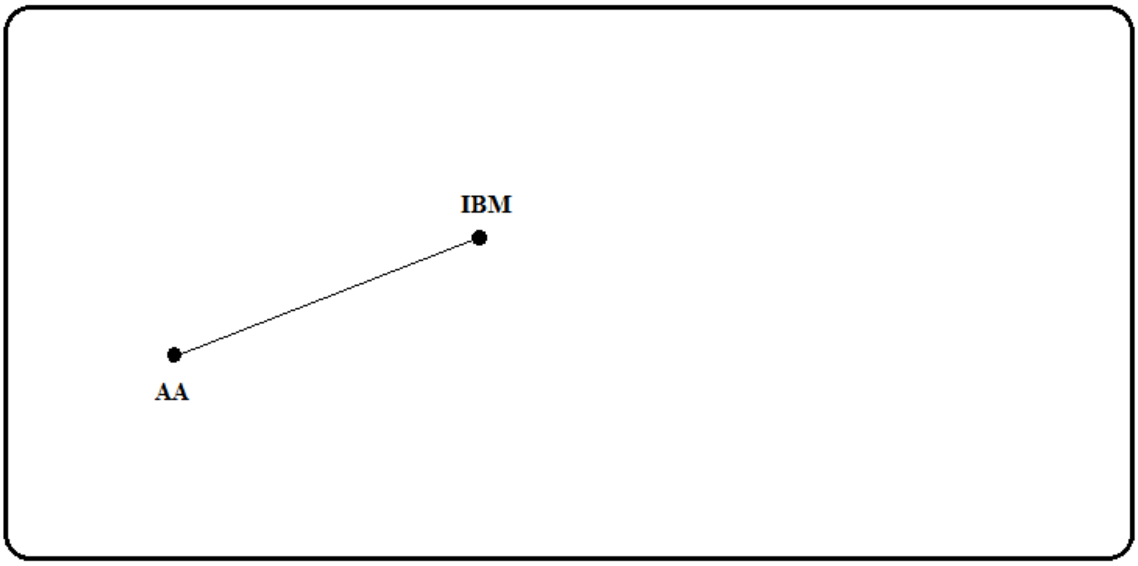


Figure 2.7. The closest two stocks are IBM and AA therefore are linked first.

	AA	XOM	AXP	BOA	IBM	JPM
AA	0	0.600	0.486	0.596	~~0.463~~	0.712
XOM		0	0.692	0.751	0.573	0.761
AXP			0	0.431	0.451	0.427
BOA				0	0.560	0.233
IBM					0	0.450
JPM						0

Table 2.8. IBM and JPM are the next closest two stocks.

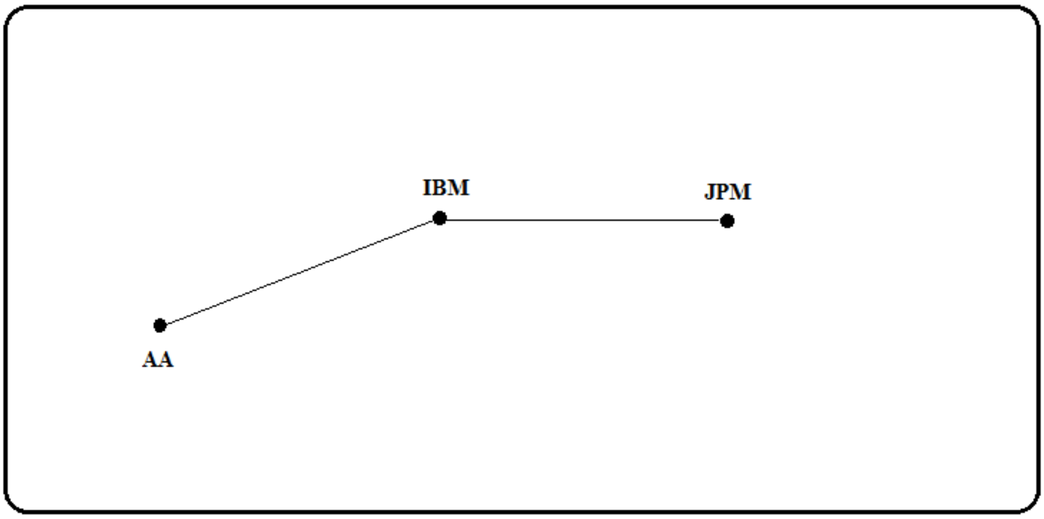


Figure 2.8. JPM are connected with IBM.

	AA	XOM	AXP	BOA	IBM	JPM
AA	0	0.600	0.486	0.596	~~0.463~~	0.712
XOM		0	0.692	0.751	0.573	0.761
AXP			0	0.431	0.451	0.427
BOA				0	0.560	0.233
IBM					0	~~0.450~~
JPM						0

Table 2.9. The next shortest distance is 0.233 between BOA and JPM.

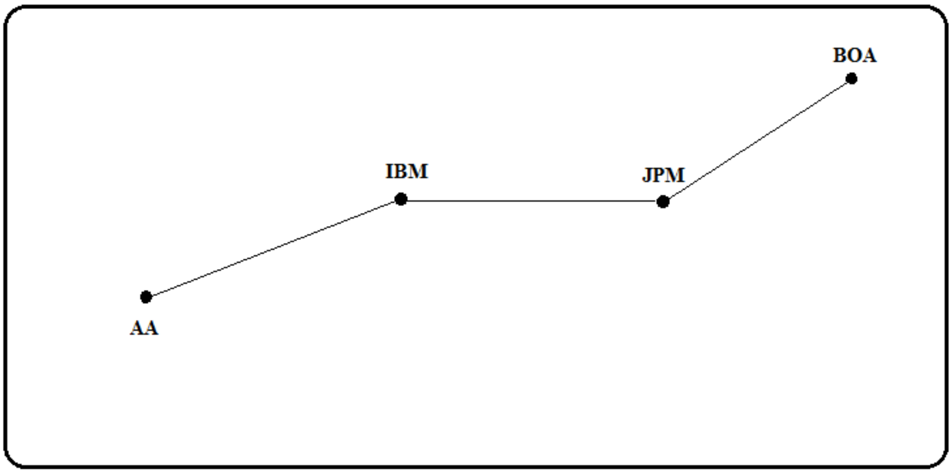


Figure 2.9. BOA is connected with JPM.

	AA	XOM	AXP	BOA	IBM	JPM
AA	0	0.600	0.486	0.596	~~0.463~~	0.712
XOM		0	0.692	0.751	0.573	0.761
AXP			0	0.431	0.451	0.427
BOA				0	0.560	~~0.233~~
IBM					0	~~0.450~~
JPM						0

Table 2.10. The next shortest distance among unconnected stocks is 0.427 between AXP and JPM.

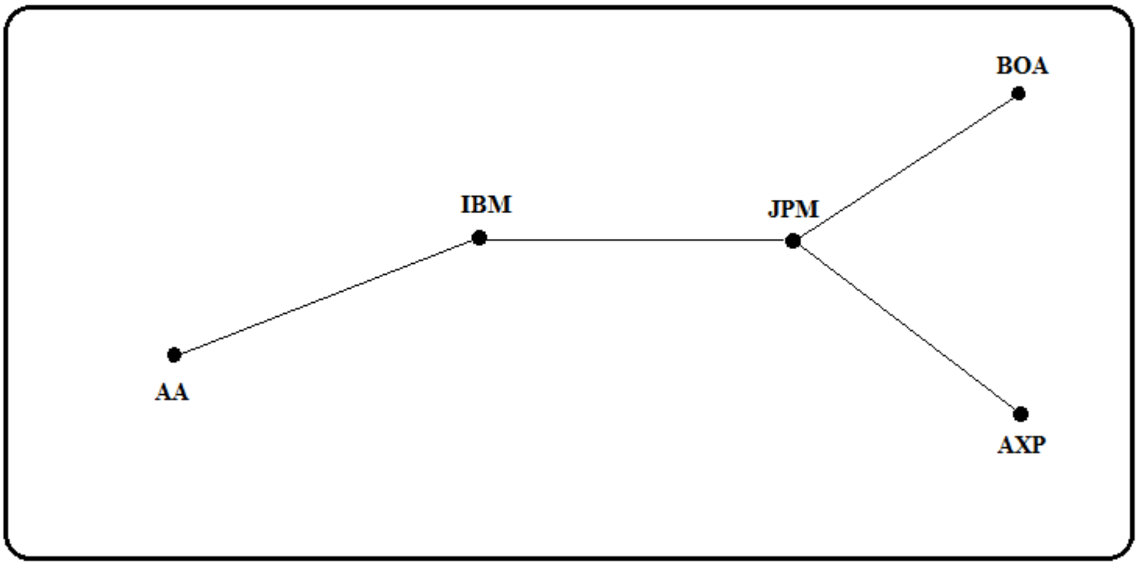


Figure 2.10. AXP is connected to JPM.

	AA	XOM	AXP	BOA	IBM	JPM
AA	0	**0.600**	0.486	0.596	0.463	0.712
XOM		**0**	**0.692**	**0.751**	**0.573**	**0.761**
AXP			0	0.431	0.451	0.427
BOA				0	0.560	0.233
IBM					0	0.450
JPM						0

Table 2.11. The last unconnected stock is XOM, it is closest to IBM.

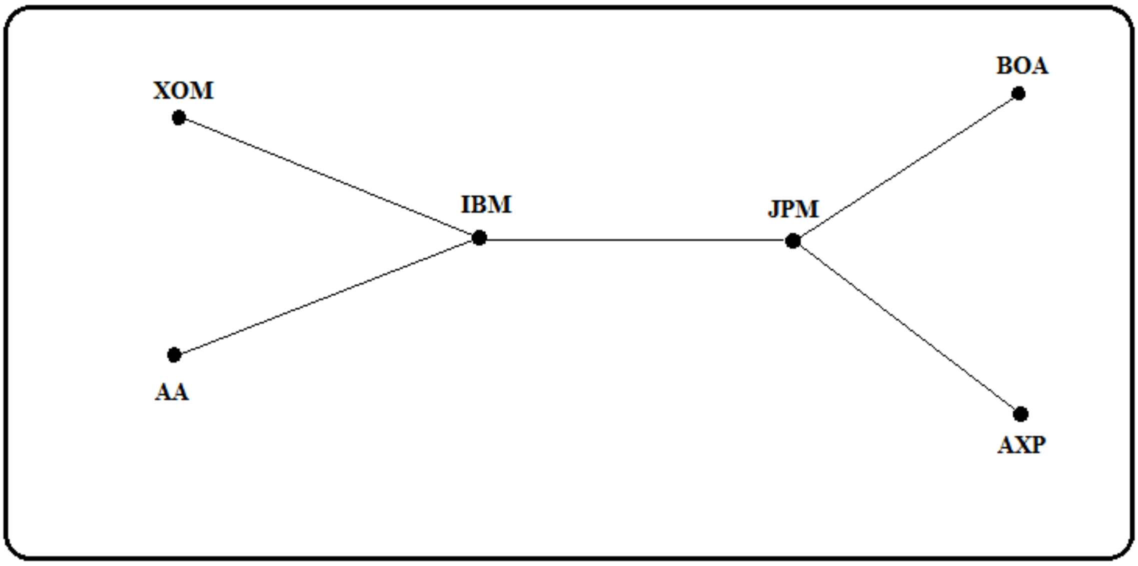


Figure 2.11. XOM is connected to IBM.

2.2 Review of portfolio diversification – theory and application

2.2.1 Modern mean-variance portfolio optimization theory

The modern portfolio theory is also known as mean-variance portfolio selection theory. It was first introduced by Harry M. Markowitz in 1952 (Markowitz 1952). His book "Portfolio Selection" (Markowitz 1959) laid out foundation for modern portfolio diversification. The theory attempts to maximize portfolio's expected return for a given amount of portfolio risk, or equivalently minimize risk for a given level of expected return by carefully choosing the proportions of various assets.

The theory was developed in the 1950s through 1970s. Since then, many theoretical and practical criticisms have been raised against it (Brodie et al., 2009; Damghani, 2013). One of the criticisms is that to apply the theory, one has to know the expected return of assets and the variance and covariance of all assets. To get assets' expected returns, ideally one needs to forecast assets' returns and covariance. For example for the ASX 200, it requires 200 expected returns, 200 variance and 19,900 $\left(\frac{200\times 199}{2}\right)$ covariances and the 19,900 covariances are not independent.

An alternative solution is to use historical data (DeMiguel et. al 2009). To estimate the variance-covariance matrix for 25 stocks with sufficient accuracy for use with the Markowitz optimization procedures, it requires 250 years of monthly returns according to DeMiguel. In reality, it becomes extremely difficult to collect that much data. If only small amount of historical data are used, that will only lead to choosing stocks with highest historical returns (Bernstein, 2001).

2.2.2 Review of other portfolio diversification methods

Given the problems associated with the mean-variance portfolio selection theory, people started to focus only on "risk-based" assets allocation strategies. Lee (Lee 2011) put these risk based approaches into the same context of mean-variance efficiency in an attempt to understand their theoretical underpinning. There were total four risk-based asset allocation methods examined in his paper: **(1). Equally Weighted Portfolio** where all assets are given the same weights. That is, for n assets, each asset will be assigned a weight equal to 1/n. Despite its simplicity, this asset allocation worked surprisingly well compared with some other common asset allocation models (DeMiguel et al. 2009). DeMiguel even concluded that "[t]here are still many miles to go before the gains promised by optimal portfolio choice can actually be realized out of sample" (DeMiguel et al. 2009, page 1915). **(2). Global Minimum – Variance Portfolio** which "*is the portfolio of risky assets that is expected to have the lowest possible volatility and that can be uniquely determined merely by a covariance matrix*" (Lee 2011, page 15). It can have higher returns and lower volatilities than the market portfolio (Clark, 2006) but it has a tendency to load up only assets that have low volatilities. **(3). Most Diversified Portfolio (MDP)** which is another version of mean-variance diversification with the "mean" portion ignored. It simply assumes all assets have the same Sharpe ratio (Sharpe 1966 and 1994) and the only objective of MDP is too maximise diversification. However, to assume all assets have the same Sharpe ratio but at the same time admit that correlations among all assets are differing from 1 implies that arbitrage opportunities exist. **(4). Risk Parity** is a special case of risk contribution portfolios; it allocates assets within a portfolio so that the risk contribution of each asset is equal within the portfolio. While having gained attention in recent years, this approach faces difficulty in finding numerical solutions. Lee (2011) concluded that risk-based asset allocation methods are the same as the mean-variance asset allocation method in the aspect that both methods require investors to have some views on assets' returns and risks.

2.3 Putting them together: diversify portfolios by visualizing correlation clusters

The idea of using visualization methods to diversify stock portfolios was initiated based on the fact that private investors hold on average only 4.3 stocks in their stock portfolios despite the advice of allocating large portion of their funds in indices and as a result they constantly underperform the market indices (Barber and Odean, 2008).

Although picking stocks from different industry groups is a viable way of diversifying portfolios, we can test whether diversification in small private investor sized portfolios can be improved by directly identifying correlation clusters. This is based on the fact that there are several levels of industry groups of available for a given stock market. The visualization methods are sufficiently flexible that they can be used to identify small numbers of correlation clusters.

This leads us to ask:

Question: Can diversification be improved in small private investor-sized portfolios using visualization to identify correlation clusters?

Chapter 3 Introduction of the new cluster visualization method – the neighbor-Net

Neighbor-Net algorithm is a distance based clustering construction algorithm developed by Bryant and Moulton (2004). The algorithm is derived from the Neighbor-Joining algorithm introduced by Saitou and Nei (1987) which was aimed at forming phylogenetic trees based on genetic data from any species, say plants and animals. Instead of constructing a tree-like phylogenetic structure, the neighbor-Net algorithm creates a network. It does this by finding the three closest neighbours that have shortest distances and combining the three neighbours into two clusters. The process continues until all taxa are put into a circular ordering structure from which a splits graph is constructed (using the splits defined by the ordering and non-negative least squares). The result of the final product is then a circular order and network rather than a tree-like structure. Because this algorithm can be used for any data which is represented by distances between objects, we will apply it to the distance based stock market correlation matrix and examine the splits graph it produces.

We use an example of five by five correlation matrix to illustrate the formation of the neighbor-Net splits graph. As shown in Table 3.1, the distances among the five stocks are calculated using the formula 1-correlation. The neighbor-Net algorithm first notes which cluster pairs (a cluster can be a single stock) are closest and then when a cluster has two "neighbours" these three clusters are collapsed into two clusters and the distance matrix updated. This process is repeated until there is only one cluster. In this example, the neighbor-Net algorithm firstly finds the closest three stocks which are JPM, BOA and AXP (Table 3.1 and Figure 3.1). Then assign the three stocks into two clusters u and v for the construction to continue. Cluster u contains stocks AXP and JPM and cluster v contains stocks JPM and BOA. After the new clusters are determined, we need to calculate a new set of pair-wise distances among all stocks in Table 3.2. As can be seen from Table 3.2, the distance between u and other stocks was calculated using the formula

$$d_{u,i} = {(d_{AXP,i} + d_{JPM,i})}/{3}$$

For example, the distance between u and IBM is calculated using the formula

$$d_{u,IBM} = (d_{AXP,IBM} + d_{JPM,IBM})/{3}$$

which is 0.3365.

The distance between u and v was calculated using the formula

$$d_{u,v} = {(d_{JPM,AXP} + d_{JPM,BOA} + d_{BOA,AXP})}/{3}$$

which is 0.3635.

The smallest distance in Table 3.2 is 0.3365, the distance between v and IBM and then there is no need to calculate any distance to position AA since it was the last stock remained and the complete neighbor-Net splits graph which is produced by software named splitstree4 (Retrieved from http://www.splitstree.org/) is in Figure 3.3.

	AXP	**JPM**	**BOA**	**IBM**	**AA**
AXP	0	**0.427**	0.4311	0.4514	0.4863
JPM		0	**0.2326**	0.4493	0.7117
BOA			0	0.5601	0.5965
IBM				0	0.463
AA					0

Table 3.1 The distances among the eight stocks calculated using the formula 1-correlation. The shortest two distances are 0.2326 between JPM and BOA and 0.4270 between JPM and AXP.

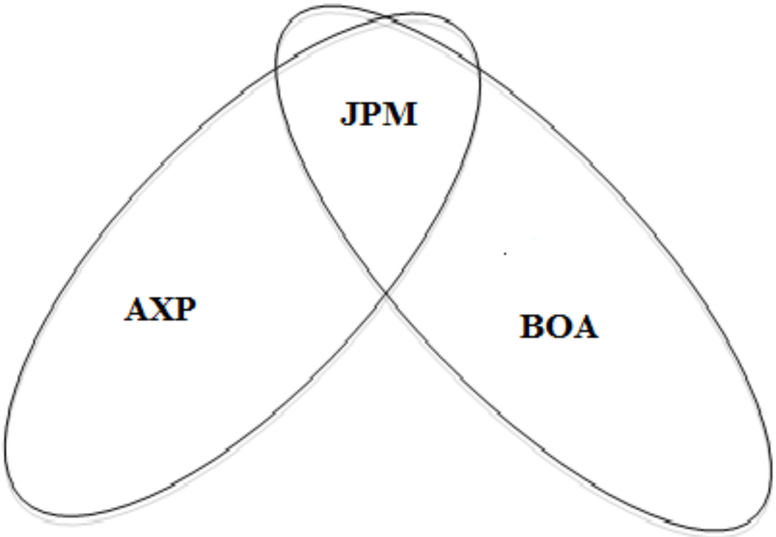


Figure 3.1 The closest three stocks are JPM, AXP and BOA.

	u	v	IBM	AA
U (AXP+JMP)	0	0.3635	**0.3371**	0.3609
V(JPM+BOA)		0	**0.3365**	0.4361
IBM			0	0.4630
AA				0

Table 3.2 "New" distances among all stocks after combining JPM, AXP and BOA. u is the combination of AXP and JPM, v is the combination of BOA and JPM. This distance between u and other single stocks is calculated using the formula

$d_{u,i} = {}^{(d_{AXP,i}\ +\ d_{JPM,i})}/_{3}$ **. For example, the distance between u and IBM is calculated using the formula** ${}^{(d_{AXP,IBM}\ +\ d_{JPM,IBM})}/_{3}$ **which is 0.3365.**

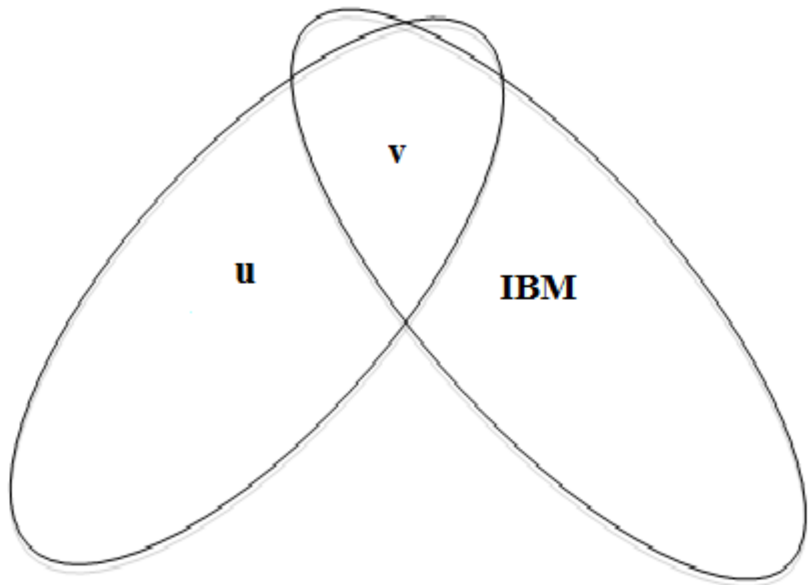


Figure 3.2 The three closest neighbours shown in Table 3.2 are u, IBM and v.

Figure 3.3 Neighbor-Net splits graph produced from "splitstree4".

Chapter 4 A note on the four different distance transformations

Correlation, which ranges from -1 to 1, is a measure of "closeness" or similarity between two time series. For instance, if the returns of two stocks have a correlation of -0.90 during a period of time, then the prices of these two stocks are regarded as "having been moving up or down opposite to each other" or "their price movements were dissimilar" during that period. On the other hand, if the returns of two stocks has a correlation of 0.90 during a period of time, then the prices of these two stocks are regarded as "having been moving up or down closely together" or "their price movements were similar" during that period.

To model the correlations using multi-dimensional scaling techniques, we need to transform the correlations to distances because negative distances do not have an obvious interpretation. To solve this problem, correlations are usually transformed so the results are non-negative, but if the absolute value is taken, it is ambiguous to determine the distance of two stocks.

Figures 4.1 to 4.4 present some popular transformations of correlations so that results are measures of dissimilarity. Mantegna (1999), Bonanno et al. (2004), Naylor et al. (2007)

$$d_{i,j} = \sqrt{2(1-\rho_{i,j})}$$

to transform correlations to non-negative since it fulfils the three distance axioms

(1) $d_{i,j} = 0$ if and only if $i=j$;

(2) $d_{i,j} = d_{j,i}$ and

(3) $d_{i,j} \leq d_{i,k} + d_{k,j}$.

However, Trosset (2005) argued that correlation is a measure of angular separation, not distance. He used the formula $d_{i,j}$= acos ($\rho_{i,j}$) to transform correlations into angles which range from 0 to π.

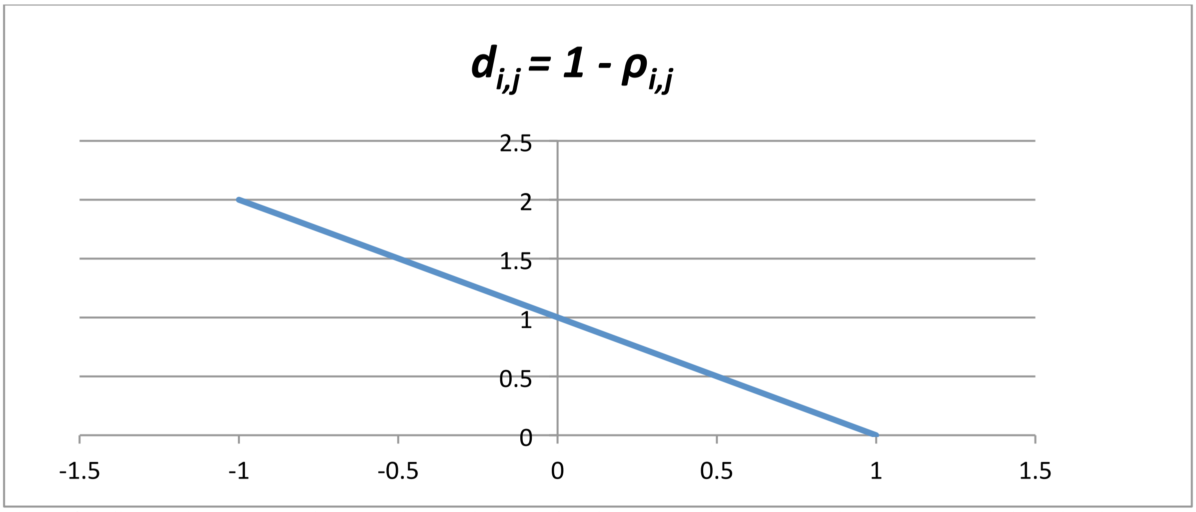


Figure 4.1. Correlations are transformed using the formula $d_{i,j} = 1 - \rho_{i,j}$. Therefore, the transformation range is from 0 to 2.

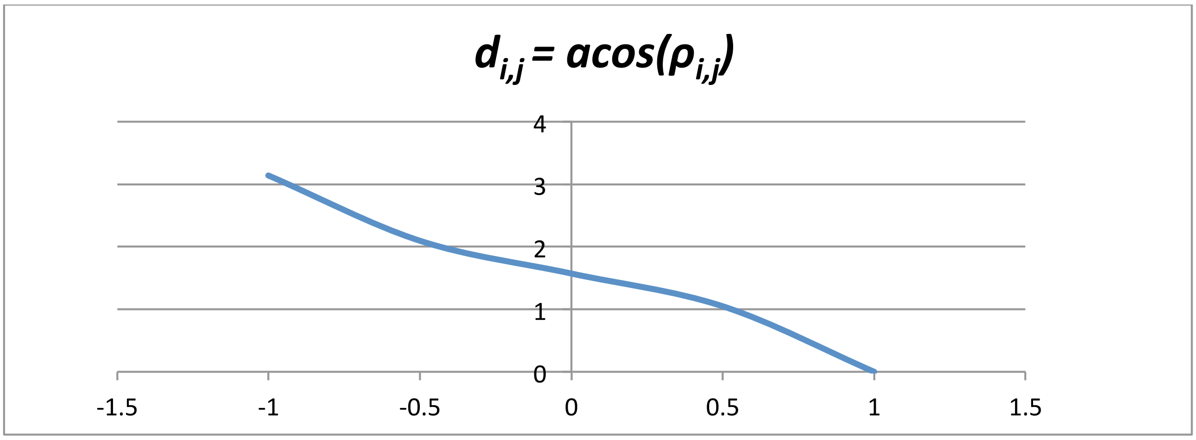


Figure 4.2. Correlations are transformed using the formula $d_{i,j} = acos(\rho_{i,j})$. Therefore, the transformation range is from 0 to π.

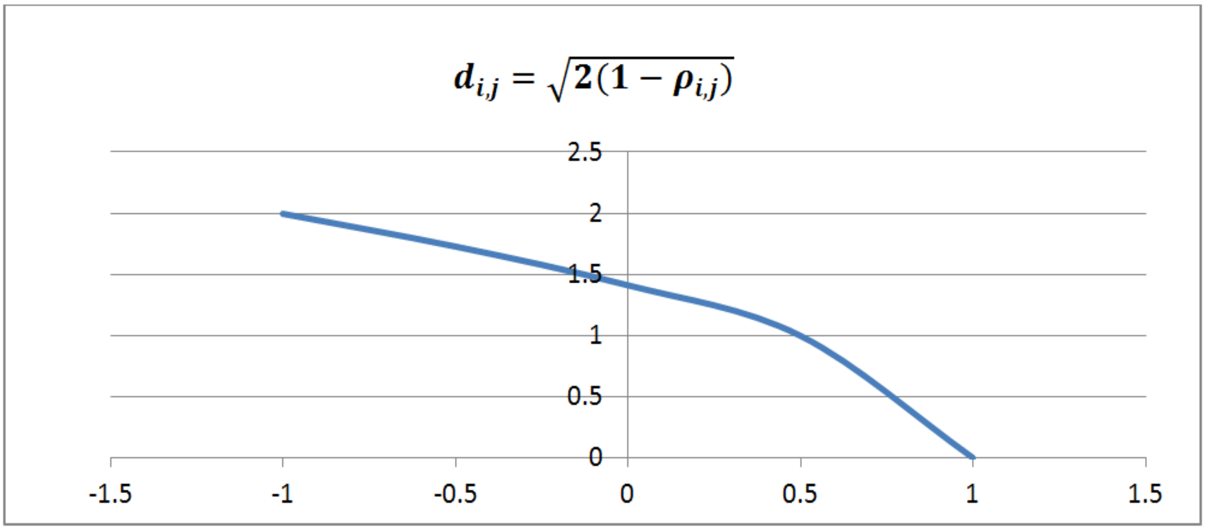


Figure 4.3. Correlations are transformed using the formula $d_{i,j} = \sqrt{2(1-\rho_{i,j})}$. Therefore, the transformation range is from 0 to 2.

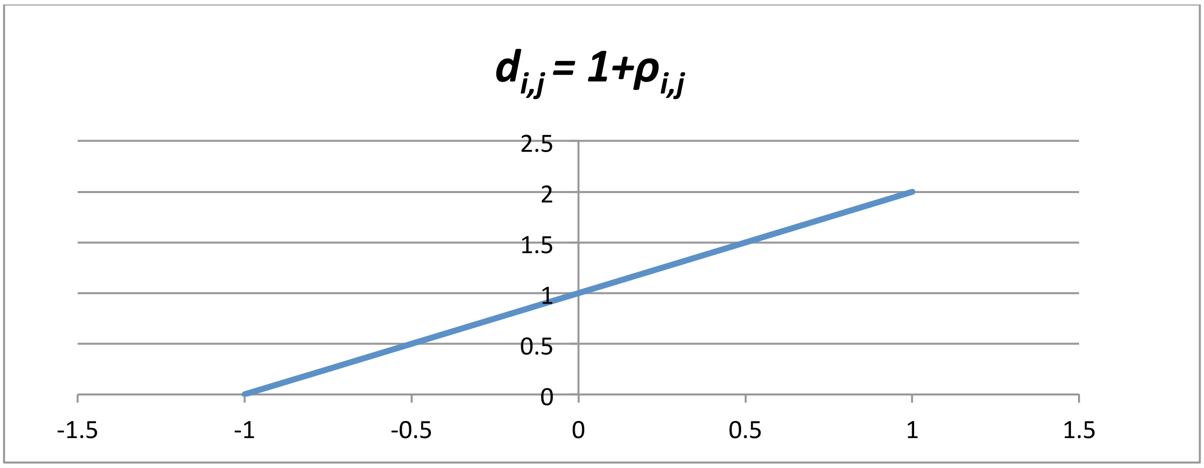


Figure 4.4. Correlations are transformed using the formula $d_{i,j} = 1+\rho_{i,j}$. Therefore, the transformation range is from 0 to 2.

In this chapter, we examine the differences among the four transformations of correlation matrix by producing HCTs, MSTs and neighbor-Net splits graphs using the four different transformations.

Figure 4.5 (a) shows the HCT produced from the correlation matrix of the 30 stocks in the Dow Jones Industrial Average for the period 20 Feb 1990 to 4 Jan 1994 using the three transformations: (1) 1- $\rho_{i,j}$, (2) acos($\rho_{i,j}$) and (3) $d_{i,j} = \sqrt{2(1-\rho_{i,j})}$. There is only one graph in Figure 4.5 (a) since the three transformations produced identical average linkage HCTs.

Figure 4.5 (b) shows the HCT produced from this period's correlation matrix after the matrix is transformed using the formula 1+ $\rho_{i,j}$.

Figure 4.6 (a) shows the MST produced from this period's correlation matrix using the three transformations. Similarly, there is only one graph in Figure 4.6 (a) since the three transformations produce identical MSTs. Figure 4.6 (b) shows the MST produced from correlation matrix after the matrix is transformed using the formula 1+ $\rho_{i,j}$.

Surprisingly, Figure 4.7 contains three different neighbor-Net splits graphs. The first neighbor-Net splits graph in Figure 4.7 (a) was produced by the correlation matrix of the period after transforming the correlation matrix using the formulas 1- $\rho_{i,j}$ and acos $(\rho_{i,j})$. Both transformations produce identical neighbor-Net splits graphs. However, the splits graph in Figure 4.7 (b), which is produced from the correlation matrix for the period after transforming the correlations matrix using the formula $\sqrt{2(1-\rho_{i,j})}$, is slightly different from the one in Figure 4.7 (a). But this difference is not large enough to split the clusters in an alternative way. In Figure 4.7 (a) and 4.7 (b) we can readily identify three clusters. The relative placement of the stocks within the cluster is slightly different but the three clusters contain the same stocks in both splits graphs. The neighbor-Net algorithm "captures" the distance based correlation matrix with more detail than that of HCT and MST.

Figure 4.7 (c) shows the neighbor-Net split graph produced from the correlation matrix after the matrix is transformed using the formula 1+ $\rho_{i,j}$. The transformed correlations become measures of similarity. Figure 4.5 (a) indicates XOM and CVX are the last two stocks to be joined to form a single cluster of all stocks, so are the most different stocks from the remaining 28. As can be seen from Figure 4.5 (b), while Figure 4.6 (b) do indicate that CVX and XOM form two hubs which indicate that they are the most different among all stocks (confirmed by Figure 4.5 (a)). Although some features of visualization using 1+ $\rho_{i,j}$ can be understood in conjunction with at least one other transformation, in general it is very hard to interpret the graphs produced by correlation matrix transformed by 1+ $\rho_{i,j}$.

In the rest of this thesis, we will use the formula $\sqrt{2(1-\rho_{i,j})}$ to transform the correlation matrices since it is the most commonly used distance transformation formula in the finance literature.

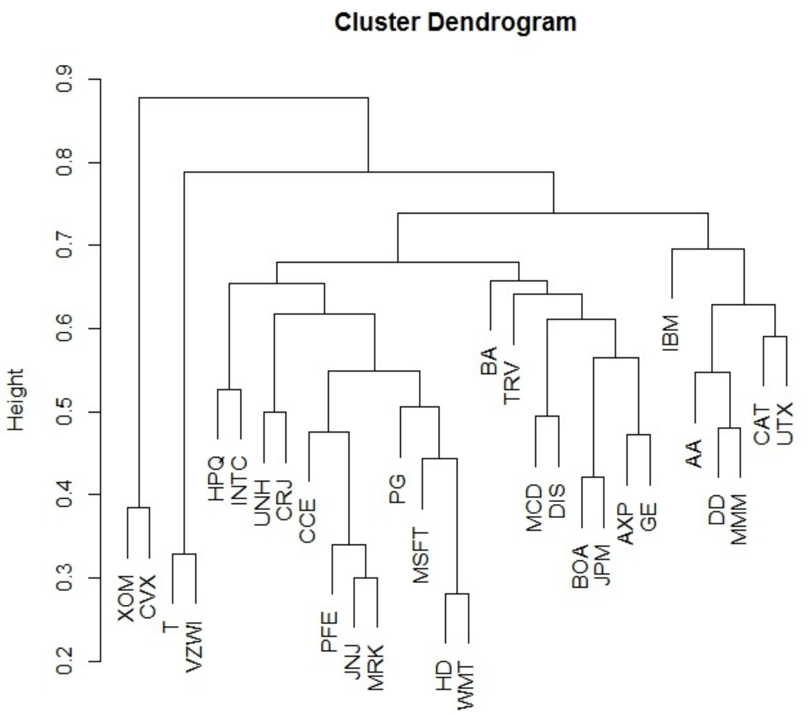


Figure 4.5 (a). Three different transformations namely 1-$\rho_{i,j}$, acos ($\rho_{i,j}$) and $\sqrt{2(1-\rho_{i,j})}$ of stock correlation for the period 20 Feb 1990 to 4 Jan 1994 produce identical single linkage HCT.

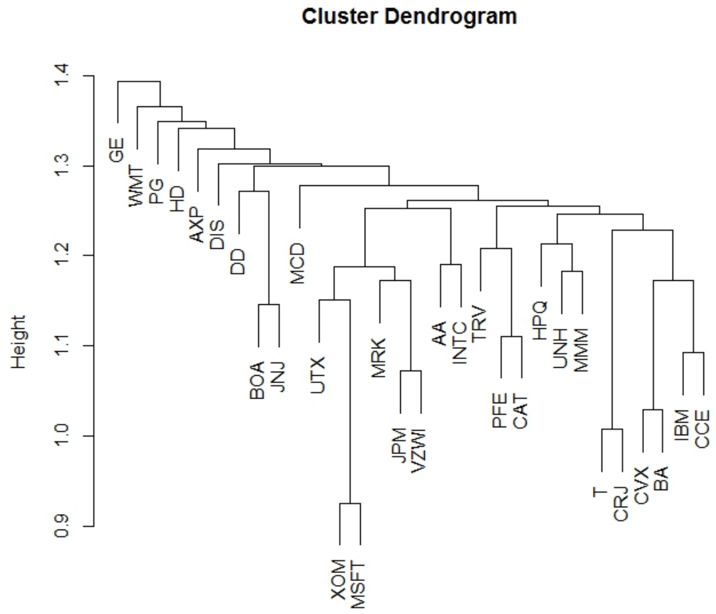


Figure 4.5 (b). The single linkage HCT produced from the correlation matrix for the period 20 Feb 1990 to 4 Jan 1994 after the correlation matrix is transformed using the formula 1+$\rho_{i,j}$.

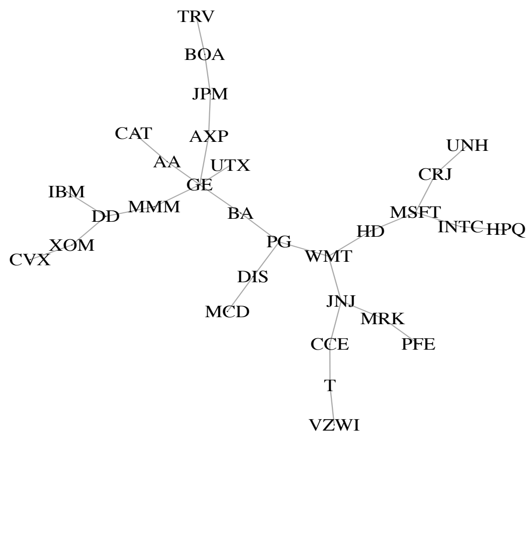


Figure 4.6 (a) Three different transformations namely 1-$\rho_{i,j}$, acos ($\rho_{i,j}$) and $\sqrt{2(1-\rho_{i,j})}$ of the correlation matrix for the period 20 Feb 1990 to 4 Jan 1994 produce identical MST.

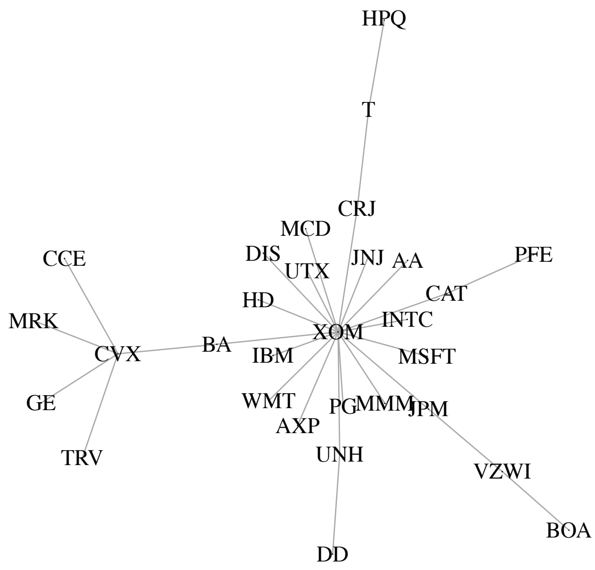


Figure 4.6 (b) MST produced from the correlation matrix for the period 20 Feb 1990 to 4 Jan 1994 after it was transformed using the formula 1+$\rho_{i,j}$.

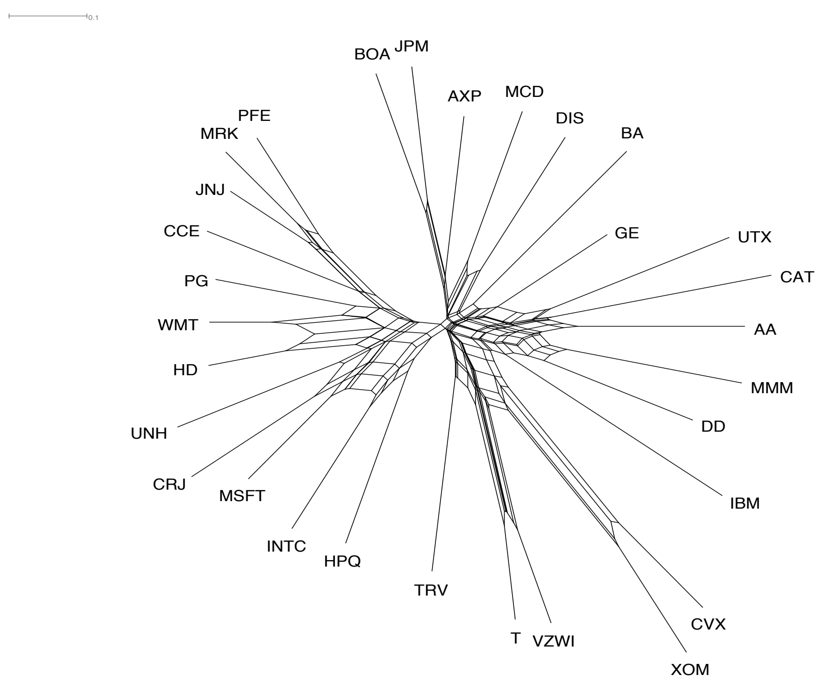


Figure 4.7 (a). Two different transformations namely 1-$\rho_{i,j}$ and acos ($\rho_{i,j}$) of the correlation matrix for the period 20 Feb 1990 to 4 Jan 1994 produce identical neighbor-Net splits orderings.

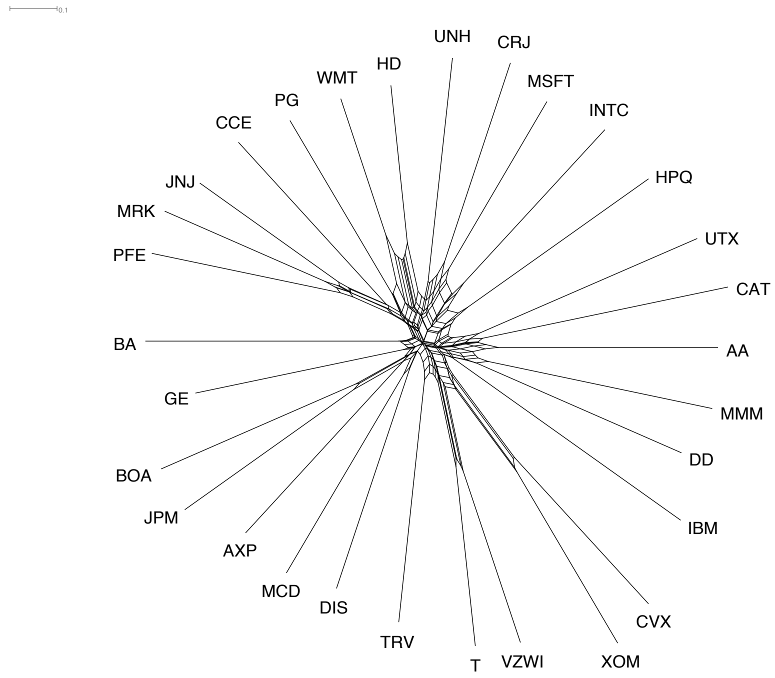


Figure 4.7 (b) The transformations $\sqrt{2(1-\rho_{i,j})}$ of the correlation matrix for the period 20 Feb 1990 to 4 Jan 1994 produce a different neighbor-Net splits ordering.

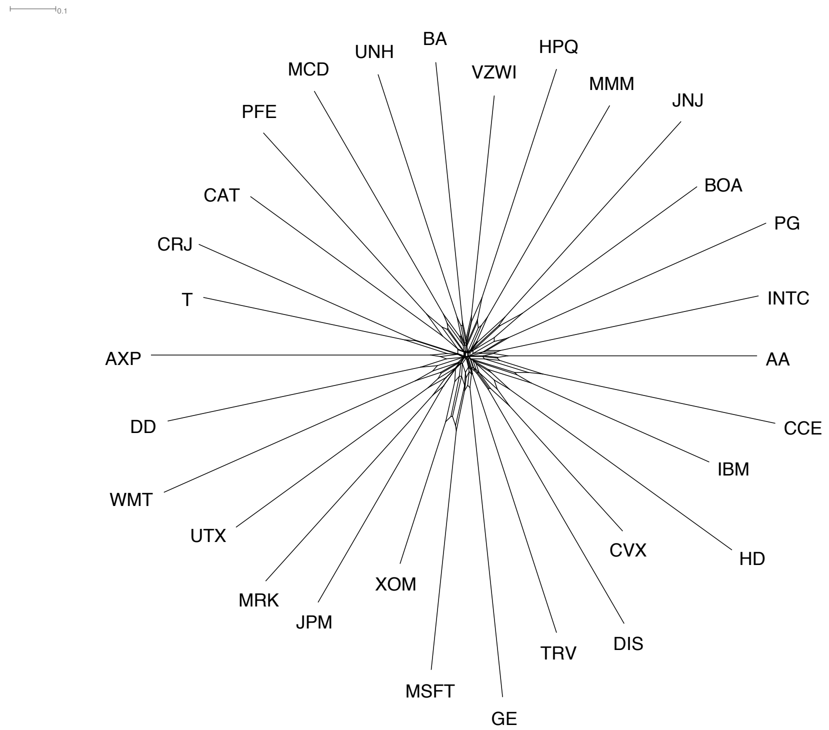


Figure 4.7 (c) Neighbor-Net split graph produced from the correlation matrix for the period 20 Feb 1990 to 4 Jan 1994 after the correlation matrix is transformed using the formula $1+\rho_{i,j}$.

Chapter 5 An initial comparison of the HCT, the MST and the neighbor-Net using Dow Jones Industrial 30 stocks data

Since the detailed descriptions and properties of each of the three algorithms: the HCT, the MST and the neighbor-Net were given in Chapters 2 and 3, in this chapter we compare the efficiencies of the three visualization methods in forming small diversified portfolios by simulating 1000 portfolios of two, four and eight stocks which are picked from different correlation clusters suggested by these three visualization methods namely the HCT, the MST and the neighbor-Net splits graph. To make the comparison comprehensive, we also added to the comparison the simulated portfolios that are picked (1) randomly and (2) from different industry groups. The rationale behind the simulation method is: if any of the visualization methods split correlation clusters more accurately[1] than the other visualization methods, then the set of 1000 portfolios formed by picking stocks from different clusters indicated by this method may have the smallest variance. The simulation process will be repeated for portfolios for five periods (Figure 5.1).

The comparison in this chapter is only exploratory; we used the Dow Jones Industrial Average which contains only 30 stocks as our database because with a small number of stocks it is easier to observe the structure of the graphs. In Chapter 6, we will examine the neighbor-Net method on a larger stock market using stocks in ASX 200. The structure of this Chapter is as follows: Section 5.1 describes the process of collecting data, transforming data and producing distance based stock market correlation matrices ready for visualization simulation. Section 5.2 explains the simulation method including the concept behind the simulation method, how the simulation works, and the motivation of using the simulation method based on a current market phenomenon, the stages of the simulation process at a

[1] Accurate implies that stocks in the same correlation clusters have high correlations, clusters further away from

glance and the actual implementation of the simulation. Section 5.3 discusses the simulation results, evaluates the performance of each method and gives our conclusions.

5.1Data

5.1.1 Data collection and transformation

As mentioned earlier, we used the stocks in the Dow Jones Industrial Average, which contains 30 stocks, as our data base. Weekly closing prices along with dividend rate and dividend date of each of the 30 stocks for the period 20 February 1990 to 14 May 2013 were collected from DataStream. The names and the industry groups of the 30 stocks, and their stock ticker symbols are listed in Appendix 1.

We divide the whole period into six shorter periods shown in Figure 5.1 and use the "out-of-sample" testing method to test the efficiencies of the three visualization methods along with the two extra methods; namely picking stocks randomly and from different industry groups in portfolio formation. That is, we use stocks' weekly return data in period 1 to create (1) Hierarchical cluster tree, (2) Minimum spanning tree and (3) Neighbor-Net splits graph, then observed period two's return distributions of portfolios' which are picked from different correlation clusters suggested by period one's graphs. Since we used out-of-sample testing, the simulation was then repeated for period three, four, five and six based on the graphs produced from the weekly returns in period two, three, four and five respectively.

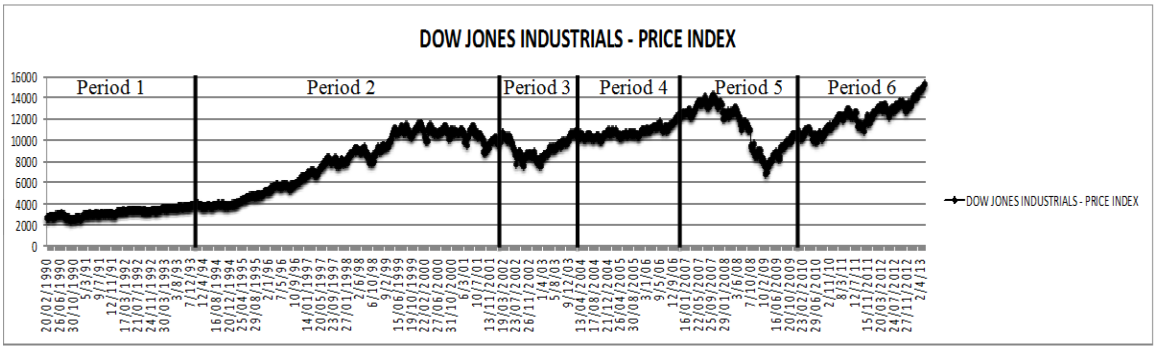


Figure 5.1. The whole period was divided into 6 shorter periods. Period 1: 20/02/1990 to 04/01/1994; Period 2: 11/01/1994 to 01/01/2002; Period 3: 08/01/2002 to 06/01/2004; Period 4: 13/01/2004 to 02/01/2007; Period 5: 9/01/2007 to 05/01/2010 and Period 6: 12/01/2010 to 14/05/2013.

Therefore, three measurements were needed to carry out the simulation:

1. Individual stock's weekly returns for periods one to five which were used to calculate stock the correlation matrices for each period. The individual stock's weekly returns for period one to five were calculated using the formula $\frac{p_{i+1}-p_i}{p_i} \times 100\%$, p_i was the closing weekly price for week i. The dividends were omitted from the calculation based on the simplifying assumption that periodic dividend payments have very little effect on stock correlations.

2. Individual stock's return for periods two to six which are calculated using the formula $\frac{p_{end} \times (accumulated\ dividend\ rate) - p_{beginning}}{p_{beginning}} \times 100\%$. Appendix 2 shows the returns for each stock for periods two through six. p_{end} is the price of the stocks at the end of the periods, $p_{beginning}$ is the price of the stocks at the beginning of the periods. The accumulated dividend rates were calculated based on the assumption that all dividends were reinvested into the related stocks immediately and were kept in the portfolios till the end of the periods.

3. Individual stock's standard deviation of weekly returns for period two through six using the formula $\sigma_i = \sqrt{\mathrm{E}[{p_i}^2] - (\mathrm{E}[\bar{p}_i])^2}$ where p_i is weekly price for week i. Appendix 3 shows the weekly standard deviations of each stock's returns for periods two through six. Where p_i is the weekly price of the stocks and $\bar{p}_i$ is the average price of the period.

5.1.2 Create correlation matrices

We use the 30 stocks' weekly returns for periods one to five to create five correlation matrices in the software package R (retrieved from http://www.r-project.org/). Appendix 4 shows the correlation matrices of stocks' weekly returns for periods one as an example.

5.1.3 Transform correlation matrices to distance matrices

Correlations in each matrix were transformed into distance measures using the formula $\sqrt{2(1-\rho_{i,j}})$. After the transformation, the new "correlations" in the matrices ranged from 0 to 2. (Also refer to Chapter 4)

5.2 Method

5.2.1 The concept and motivation behind the simulation method and the illustration of the method

5.2.1.1 The concept behind the simulation method

One of the objectives of mean-variance portfolio theory is to minimise variance of portfolios for a given level of return (Markowitz, 1952). The variance of a two stock portfolio can be written as: $\sigma_p^2 = \sum_i w_i^2 \sigma_i^2 + \sum_i \sum_{j \neq i} w_i w_j \sigma_i \sigma_j \rho_{ij}$, where ρ_{ij} is the correlation coefficient between the returns on assets i and j. The portfolio's return volatility (standard deviation) is then $\sigma_p = \sqrt{\sigma_p^2}$.

In modern portfolio diversification, the general strategy of forming portfolios is to pick stocks that are not highly correlated or in other words weakly correlated. One popular pragmatic strategy is to pick stocks from different industry groups. Because stocks within the same industry group are believed to be highly correlated in price movements, stocks from different industry groups are believed to have low correlation in price movements.

Therefore, variance of portfolio of two stocks that are picked from two different industry groups is:

$$\sigma_p^2 = \sum_{iJ} \sum_{gK} w_{iJ}\, w_{gK} \sigma_{iJ} \sigma_{gK} \rho_{iJ,gK},$$

where i and g are individual stocks in industry group J and K (see Figure 5.2).
Similarly, the variance of portfolio of two stocks that are picked from the same industry group (industry group K) is:

$$\sigma_p^2 = \sum_{iK} \sum_{gK} w_{iK}\, w_{gK} \sigma_{iK} \sigma_{gK} \rho_{iK,gK},$$

Where i and g are individual stocks in industry group K.

Since $\rho_{iJ,gK} < \rho_{iK,gK}$, the portfolios of 2 stocks that are picked from different industry group are expected to have smaller variance than a portfolio picked entirely from the same industry.

The smaller the correlation of the two stocks the smaller that variance of the portfolio. Therefore to form a portfolio of two stocks, it is rational to pick stocks from two clusters with low correlation. In reality, some investors pick one stock from one industry group and pick the second stock from another industry group that are expected to be only weakly related to the first industry group.

As shown in Figure 5.2, stocks within one industry group are believed to be highly correlated in prices or returns. So the variance of portfolio of two stocks that were picked from two different industry groups is expected to be smaller than that of the portfolio of two stocks that were picked from the same industry group.

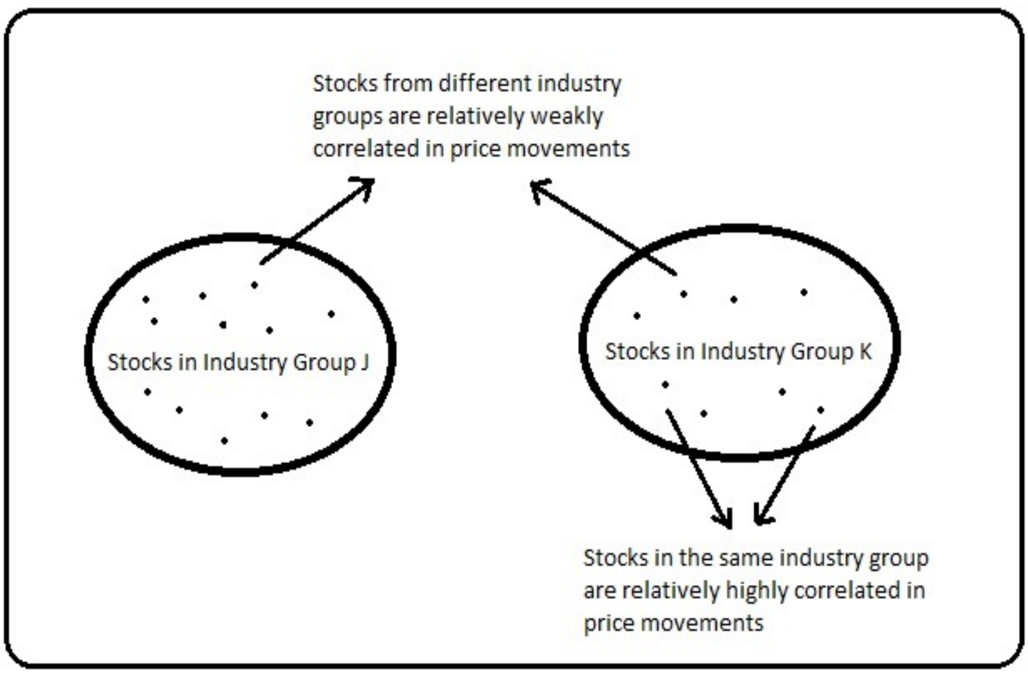


Figure 5.2 Portfolios of two stocks picked from two different industry groups are expected to have smaller variances than the portfolios of two stocks picked from the same industry group.

Definition of correlation cluster group

The correlation cluster groups are the clusters within which stocks are highly correlated in price movements (Figure 5.3). By now, we have two broad methods to recognise correlation clusters. The first one is by recognising the industry groups. In this sense, the correlation clusters are the industry groups. The second method of recognising correlation clusters is using the visualization algorithms we studied in chapters 2 and 3. Therefore, it is beneficial to compare the different ways of defining correlation clusters and we hope to find an alternative approach to portfolio selection, especially for portfolios held by small investors.

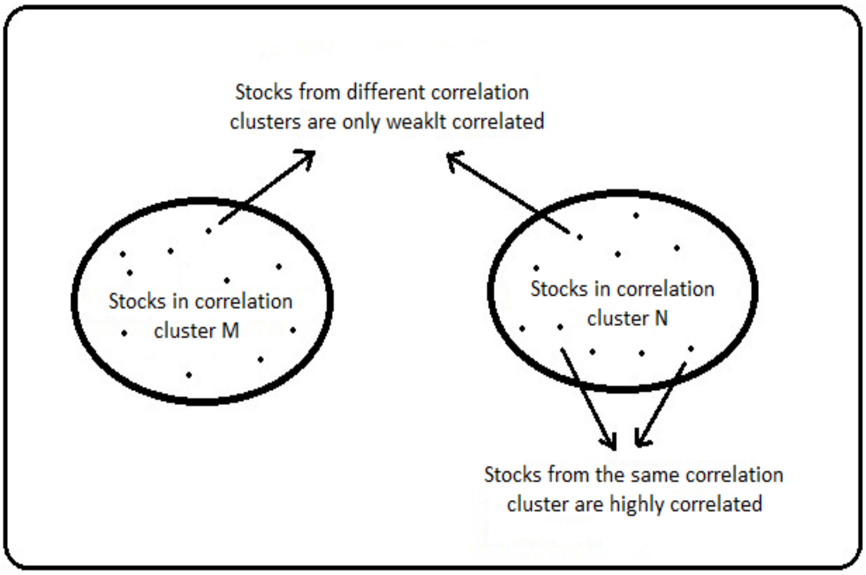


Figure 5.3 Portfolios of two stocks picked from two different correlation clusters are expected to have smaller variances than the portfolios of two stocks picked from the same correlation cluster.

5.2.1.2 Motivation for using simulation method

Although it is a viable strategy to picking stocks from different industry groups, stocks from same industry groups do not necessarily have high correlations in price movements and similarly stocks from different industry groups do not necessarily have low correlations in price movements.

Therefore it would be informative to form portfolios by picking stocks from the different correlation clusters suggested by the three algorithms. By comparing risks and returns of the portfolios using out-of-sample testing, we hope to find a better approach to portfolio diversification.

Small investors should invest a substantial amount of their funds in index funds, but in reality they do not. In fact, one of the phenomena in the area of investment funds management is that managed funds (also called actively managed funds) consistently underperform the relevant market index yet as high as 89% of the total investment funds are actively managed funds (The Economist May 3rd 2014). It would be beneficial to determine if any other diversification strategy is better than picking stocks from different industry groups. Since the average number of stocks in a private investor's portfolio in US, for example, is 4.3 (Barber and Odean, 2008), we will test if any of the visualization methods we introduced in previous chapters provides a better approach to forming portfolios that contain only a small number of stocks.

The design of the simulation is based on the fact that managed funds consistently underperform the market index. The 1000 portfolios can represent 1000 investors including fund managers and private investors who manage their own funds. We assume each investor holds either two, four or eight stocks. Then the standard deviation of the 1000 portfolios will give us an indication of the kurtosis of their underperformance. A smaller standard deviation means given an underperforming portfolio, the underperformance is likely to be less than if the standard deviation is large. We hope to find a visualization method that provides a better approach to forming portfolios that contain only small numbers of stocks.

5.2.1.3 The illustration of the simulation method

We use the open source software package R (Turner 2011) to simulate 1000 portfolios each of which is formed by picking stocks:

1. randomly
2. from the correlation clusters that are determined by the previous period's HCT
3. from the correlation clusters that are determined by the previous period's MST
4. from the correlation clusters that are determined by the previous period's neighbor-Net splits graph;
5. from different industry groups

We explain each strategy of picking stocks below:

Pick stocks randomly: Stocks were picked randomly using a uniform distribution without replacement. In other world, each stock was given equal chance of being selected but no stock was selected twice within a single portfolio.

Pick stocks from different industry groups: There were eight industry groups (see Appendix 1). The industry groups were randomly paired to make total of four big industry groups. For portfolios of two stocks, each stock was picked randomly from different groups without replacement. That is, no industry was selected twice and no stock was selected twice to form a portfolio. For portfolios of four stocks, one stock was selected from each industry group without replacement. For portfolio of eight stocks, two different stocks were selected without replacement from each of the four bigger industry groups.

Pick stocks from the correlation clusters that are determined by the previous period's HCT, MST and neighbor-Net splits graph: Because we use out-of-sample testing to test the efficiency of the visualization method at splitting clusters, we use the HCT, MST and neighbor-Net splits graph that was made from previous period's data. That is, use period one, two, three, four and five's data to produce HCT, MST and neighbor-Net splits graphs, and test them by examining the period two, three, four, five and six's portfolios returns and standard deviations. The clusters were determined by the structure of HCT, MST and neighbor-Net splits graphs which were easily observed by looking at these graphs. For simplicity, we split all graphs into four clusters in every period. For portfolios of two stocks, each stock was picked randomly from two different hierarchical clusters, MST clusters and neighbor-Net splits graphs clusters without replacement. That is, no cluster was selected twice and no stock within a cluster was selected twice for the same portfolio. For portfolios of four stocks, one stock was picked from each of the four clusters. For portfolio of eight stocks, two different stocks were selected from each of the four clusters.

Therefore for each period, we generate five return density curves and five returns to weekly volatility distribution graphs based on the five stock picking methods; they are similar to the graphs in Figure 5.4 and Figure 5.5.

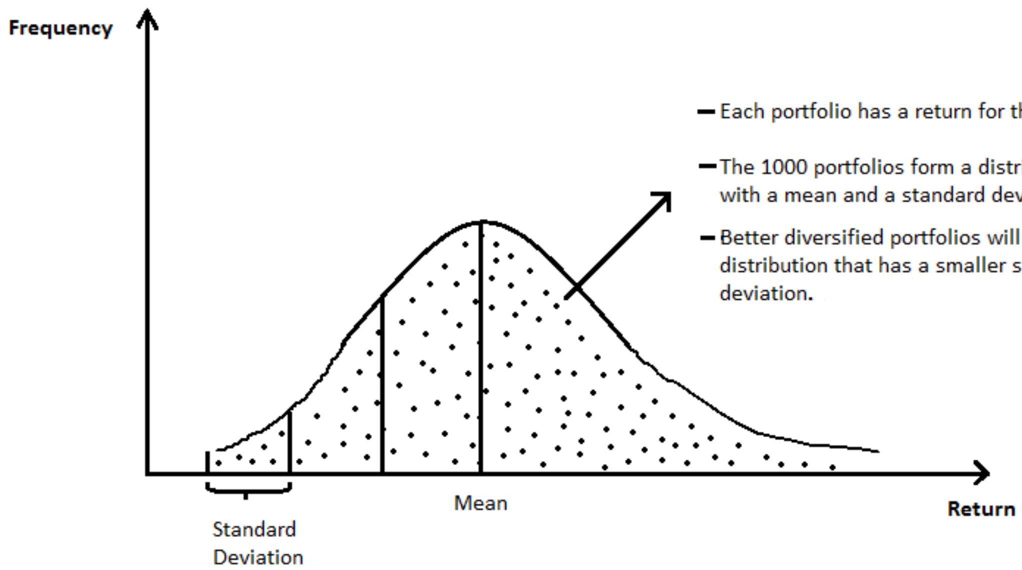


Figure 5.4. The mean and frequency of the 1000 portfolios.

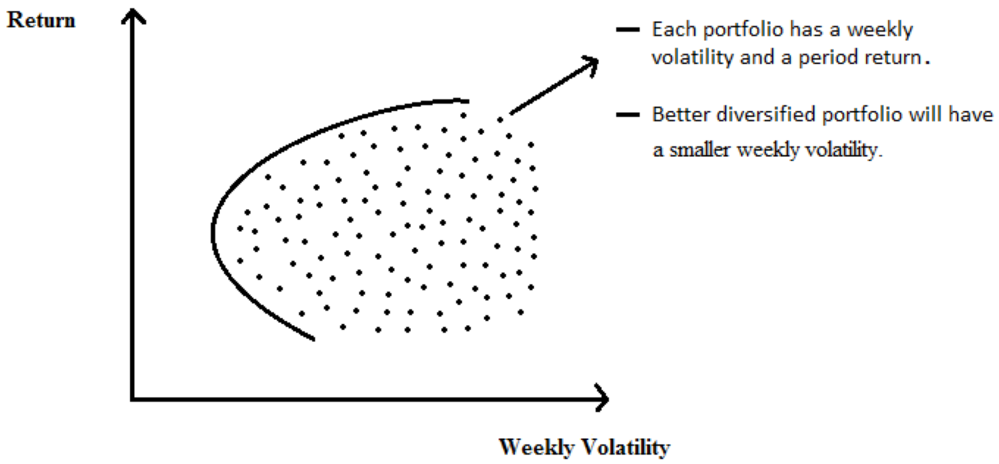


Figure 5.5. Weekly volatility and return of the 1000 simulated portfolios.

5.2.2 Stages of the simulation method at a glance

Step 1 Create return correlation matrices

- Create return correlation matrices for the 30 stocks using their weekly returns in periods one to five.

Step Transform correlation matrices into distance matrices

- Transform correlation matrices into distance matrix using the formula √2(1-ρ).

Step 3 Make the three types of graphs

- Up load the distance matrix into R and Splitstree4 to create the hierarchical cluster tree, the Minimum spanning tree and the neighbor-Net splits graph.

Step 4 Determine correlation cluster groups

- Determine correlation clusters produced by the Hierarchical cluster tree, the Minimum spanning tree and the neighbor-Net splits graph.

Step 5 Do simulations in R

- Simulate five sets of 1000 portfolios using five different strategies for period two; repeat the simultion process for period three, four, five and six.

Figure 5.6. Steps of the simulation process.

5.2.3 Implementation of the simulation method

5.2.3.1 Create the HCTs, the MSTs and the neighbor-Net splits graphs for period one through five

Upload the matrices into R and splitsTree4 (Retrieved from http://www.splitstree.org/ & Huson, 2006). Use R to create the Hierarchical cluster trees, the Minimum spanning trees and use the splitstree4 to produce the neighbor-Net splits graph.

Finally, the period 1's graphs and clusters are in Figure 5.7 (a), 5.8 (a) and 5.9 (a) and in Table 5.1 (a), (b) and (c). The three types of graphs for period two, three, four and five are in Appendix 6. Each graph is split into four clusters. Note that companies' names are the one to four letter stock ticker symbols for convenience. The companies and their codes are listed in Appendix 1.

Initial observation of the graphs:

The initial observation shows that the clusters suggested by the three methods are different. Tables 5.1 (a), (b), (c) (in this chapter) and Table 5.2 (a) to 5.5 (c) (in Appendix 6) listed the stocks in each of the four clusters suggested by each of the three algorithms. The clusters split by the neighbor-Net splits graphs were observed by eye, that is, the stocks that were close to each other were grouped within in the same clusters. The clusters from the hierarchical cluster tree were determined by the split shown by the hierarchical cluster trees. One important property of the clusters split by the HCT is that the clusters are highly unbalanced. As shown in Figure 5.8 (a) and Table 5.1 (b), there are two clusters that contain only two stocks and another cluster contains 20 stocks. This highly unbalanced structure will produce artificially low standard deviation and hence high Sharpe ratios. It is unrealistic to split clusters in such way in reality. The drawback of the unbalanced splits will be discusses in the results section. The clusters of the MST were also determined by eye.

5.2.3.2 Determine correlation clusters

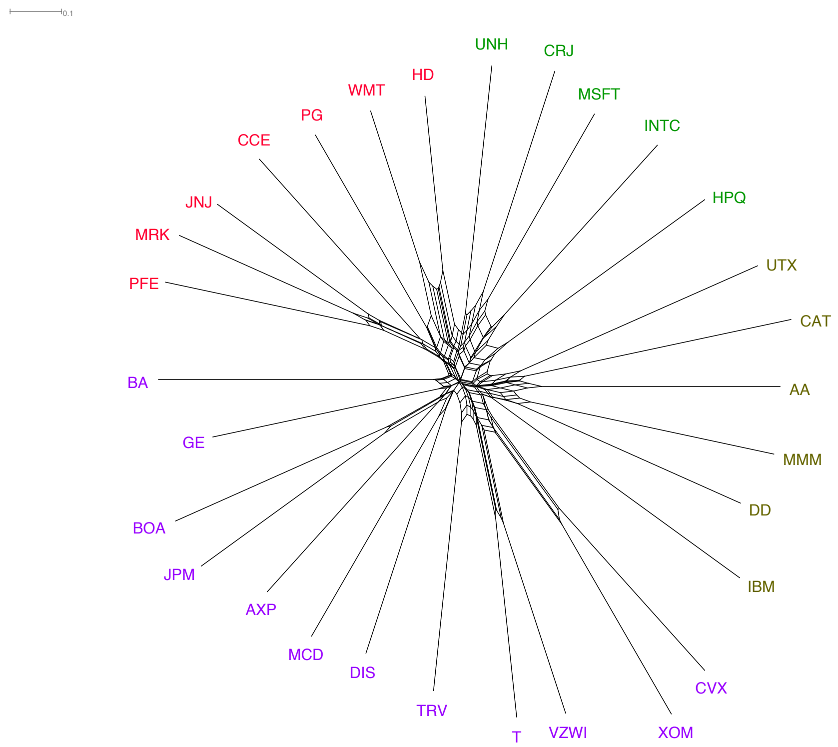


Figure 5.7 (a) Neighbor-Net splits graph produced from stocks' weekly returns in period 1: 20 February 1990 to 04 January 1994.

Cluster 1 (in Red)	PFE,MRK,JNJ,CCE,PG,WMT,HD
Cluster 2 (purple)	BA,GE,BOA,JPM,AXP,MCD,DIS,TRV,T,VZWI,XOM,CVX
Cluster 3 (Dark Green)	UTX, CAT, AA, MMM, DD, IBM
Cluster 4 (Green)	UNH,CRJ,MSFT,INTC,HPQ

Table 5.1 (a). The 4 clusters determined by the neighbor-Net splits graph of period 1: 20 February 1990 to 04 January 1994.

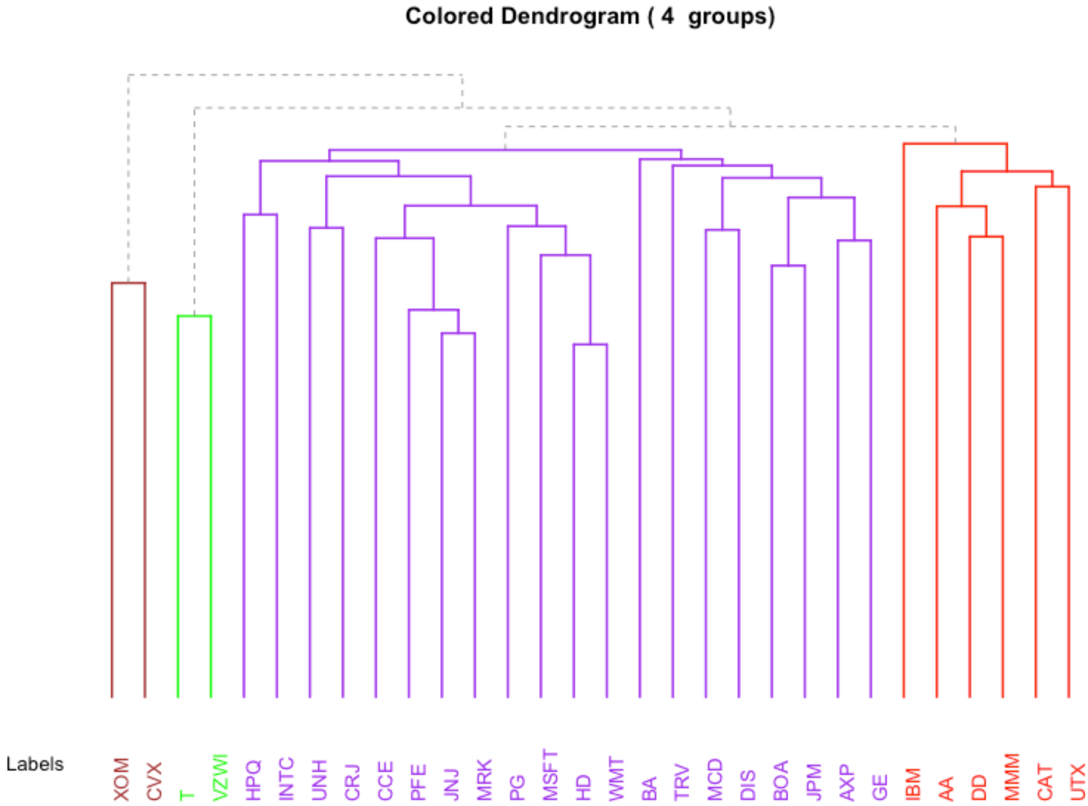


Figure 5.8 (a) The Hierarchical cluster produced from stocks' weekly returns in period 1: 20 February 1990 to 04 January 1994.

Cluster 1 (in brown)	XOM, CVX
Cluster 2 (in green)	T, VZWI
Cluster 3 (in purple)	HPQ, INTC, UNH, CRJ, CCE, PFE, JNJ, MRK, PG, MSFT, HD, WMT, BA, TRV, MCD, DIS, BOA, JPM, AXP, GE
Cluster 4 (in Red)	IBM, AA,DD,MMM,CAT,UTX

Table 5.1 (b). The 4 clusters determined by the HCT of period 1: 20 February 1990 to 04 January 1994.

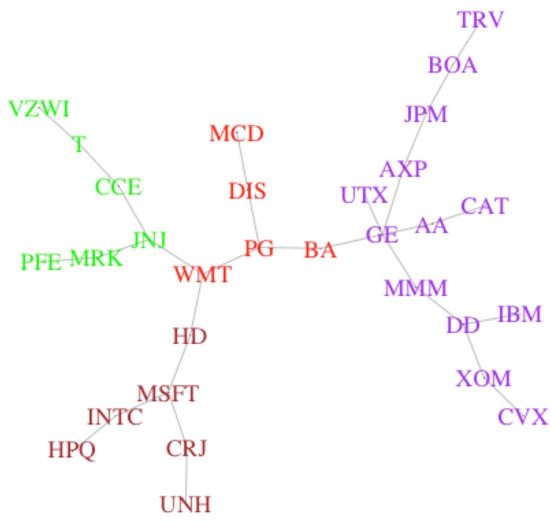


Figure 5.9 (a) MST produced from the stocks' weekly returns in period 1: 20 February 1990 to 04 January 1994.

Cluster 1(in Red)	BA, PG, WMT, DIS, MCD
Cluster 2 (purple)	CVX, XOM, IBM, DD, MMM, CAT, AA, GE, UTX, AXP, JPM, BOA, TRV
Cluster 3 (in brown)	HD, MSFT, CRJ, UNH, INTC, HPQ
Cluster 4 (in green)	VZWI, T, CCE, JNJ, MRK, PFE

Table 5.1 (c). The 4 clusters determined by the MST of period 1: 20 February 1990 to 04 January 1994.

5.2.3.3 Run simulations in R

The simulations were run in R. R code is available upon request from

cheng.zhan@pg.canterbury.ac.nz

5.3 Result and discussion

The simulation results including the means, standard deviations and Sharpe Ratios of each set of simulated return are presented in Tables 5.6 through 5.10.

Two types of graphs were also created and presented in Appendix 6. The first type is the histogram and probability curves of the simulated portfolios' period returns and frequencies of returns. The second type of graph is the weekly variance and returns scatter plot.

Although there no single method of diversification consistently outperformed the other methods, there are some significant and unexpected results shown in the tables. We highlight and discuss them in turns as follow:

(a) HCT methods produced the highest Sharpe ratio for periods two, four, five and six and for all sizes of portfolios except for the eight-stock portfolios in period six where HCT produced the second highest Sharpe ratio. As mentioned earlier, since the HCT split the clusters in a way that is highly unbalanced, it has the potential to produce artificially low standard deviations and hence high Sharpe Ratios. In reality, we would not split stocks into clusters in such an unbalanced way, therefore in next chapter we will not study this method.

(b) The industry group method did not perform better than any other method. In period three, it even produced the lowest Sharpe ratio, that is, it performed worse than the random selection methods.

(c) The neighbor-Net splits graph method did not show a consistent performance across all five periods. In period two, the two-stock portfolios produced the second highest Sharpe ratio; in period three, the two and four – stock portfolio produced the second highest Sharpe ratios; in period four, all sized portfolios produced the second highest

Sharpe ratios; In period six, both two and four – stock portfolios produced the highest Sharpe ratios and the eight – stock portfolios produced the second highest Sharpe ratio. However, in period five, it produced the lowest Sharpe ratio – lower than the random selection method. Taking into account all methods and all results and given the fact that the HTC produces highly unbalanced clusters, it is appropriate to pick the Neighbor-Net as the one method we will use for determining correlation clusters while conduct the ASX200 study.

As can be seen following each result table, we also calculated the p values of LEVENE's tests and the AVOVA tests. The p values of the AVOVA tests were very small which indicate that the means of all the portfolios created by the different selection methods are significantly different. Therefore, we need to be cautious about over-interpreting the differences in means. There is no reason to expect that they should be different.

The LEVENE's test, which simultaneously tests all standard deviations for equality, is also important because the p values show that the standard deviations of the portfolios created by the different selection methods are very different. Even though we did not do pair wise tests, we can still conclude that it is beneficial to test if directly identifying correlation clusters will reduce risk.

Period 2	Random	Neighbor-Net splits graph	Hierarchical Cluster Tree	Minimum Spanning Tree	Industry Groups
Mean return					
(2-stock portfolios)	520.89	537.15	458.26	479.12	513.45
(4-stock portfolios)	519.89	541.26	458.79	490.71	525.23
(8-stock portfolios)	514.73	541.14	455.44	483.21	522.32
Standard Deviation					
(2-stock portfolios)	175.02	174.60	132.90	158.69	170.08
(4-stock portfolios)	121.30	120.10	89.07	111.87	105.35
(8-stock portfolios)	78.51	78.83	58.82	70.06	68.70
Sharpe Ratio					
(2-stock portfolios)	**2.95**	**3.05 (2nd highest)**	**3.41 (Highest)**	**2.99**	**2.99**
(4-stock portfolios)	**4.24**	**4.46**	**5.09 (Highest)**	**4.34**	**4.94 (2nd highest)**
(8-stock portfolios)	**6.49**	**6.80**	**7.65 (Highest)**	**6.82**	**7.53 (2nd highest)**

Table 5.6. Mean returns, standard deviations and Sharpe ratios under the five different portfolio selection methods for period 2 using period 1's data for estimation of correlation clusters. Portfolios' size was two, four and eight respectively. The risk free rate used to calculate Sharpe ratios was 5.3%.

	2 - stock portfolios	4 – stock portfolios	8 – stock portfolios
AVOVA Test's P value	<0.001	<0.001	<0.001
Levene's Test's P value	<0.001	<0.001	<0.001

The p values of Levene's test and the AVOVA test for period 2.

Period 3	Random	Neighbor-Net Splits graph	Hierarchical Cluster Tree	Minimum Spanning Tree	Industry Groups
Mean return					
(2-stock portfolios)	21.91	21.52	17.86	27.45	19.78
(4-stock portfolios)	22.34	21.24	19.86	25.67	20.56
(8-stock portfolios)	22.60	21.49	19.82	25.99	19.75
Standard Deviation					
(2-stock portfolios)	20.92	20.23	18.84	17.85	21.91
(4-stock portfolios)	14.16	13.17	12.45	11.93	14.35
(8-stock portfolios)	9.38	8.81	7.30	7.49	8.92
Sharpe Ratio					
(2-stock portfolios)	**0.90**	**0.92 (2nd highest)**	**0.79**	**1.37 (Highest)**	**0.77**
(4-stock portfolios)	**1.37**	**1.38 (2nd highest)**	**1.35**	**1.90 (Highest)**	**1.22**
(8-stock portfolios)	**2.09**	**2.10**	**2.30 (2nd highest)**	**3.07 (Highest)**	**1.88**

Table 5.7. Mean returns, standard deviations and Sharpe ratios under the five different portfolio selection methods for period 3 using period 2's data for estimation of correlation clusters. Portfolios' size was two, four and eight respectively. The risk free rate used to calculate Sharpe ratios was 3%.

	2 - stock portfolios	4 – stock portfolios	8 – stock portfolios
AVOVA Test's P value	<0.001	<0.001	<0.001
Levene's Test's P value	<0.001	<0.001	<0.001

The p values of Levene's test and the AVOVA test for period 3.

Period 4	Random	Neighbor-Net splits graph	Hierarchical Cluster Tree	Minimum Spanning Tree	Industry Groups
Mean return					
(2-stock portfolios)	60.80	66.76	86.29	60.44	61.09
(4-stock portfolios)	61.46	66.75	87.44	60.95	60.14
(8-stock portfolios)	62.64	67.08	87.31	60.88	61.55
Standard Deviation					
(2-stock portfolios)	30.57	31.48	29.40	31.91	34.21
(4-stock portfolios)	21.52	20.77	18.58	21.85	22.83
(8-stock portfolios)	13.99	14.09	9.87	12.89	14.57
Sharpe Ratio					
(2-stock portfolios)	**1.92**	**2.05 (2nd highest)**	**2.86 (Highest)**	**1.83**	**1.72**
(4-stock portfolios)	**2.75**	**3.11 (2nd highest)**	**4.59 (Highest)**	**2.69**	**2.54**
(8-stock portfolios)	**4.32**	**4.60 (2nd highest)**	**8.62 (Highest)**	**4.55**	**4.07**

Table 5.8. Mean returns, standard deviations and Sharpe ratios under the five different portfolio selection methods for period 4 using period 3's data for estimation of correlation clusters. Portfolios' size was two, four and eight respectively. The risk free rate used to calculate Sharpe ratios was 2.2%.

	2 - stock portfolios	4 – stock portfolios	8 – stock portfolios
AVOVA Test's P value	<0.001	<0.001	<0.001
Levene's Test's P value	<0.001	<0.001	<0.001

The p values of Levene's test and the AVOVA test for period 4.

Period 5	Random	Neighbor-Net splits graph	Hierarchical Cluster Tree	Minimum Spanning Tree	Industry Groups
Mean return					
(2-stock portfolios)	24.72	22.09	35.48	24.90	26.50
(4-stock portfolios)	24.26	20.57	36.05	24.87	26.18
(8-stock portfolios)	24.83	20.44	36.55	25.01	25.80
Standard Deviation					
(2-stock portfolios)	23.52	24.52	18.93	22.75	21.19
(4-stock portfolios)	16.07	16.10	12.42	16.02	13.83
(8-stock portfolios)	10.18	9.59	7.43	10.65	9.03
Sharpe Ratio					
(2-stock portfolios)	**0.86**	**0.72 (Lowest)**	**1.64 (Highest)**	**0.90**	**1.04 (2nd highest)**
(4-stock portfolios)	**1.24**	**1.00 (Lowest)**	**2.55 (Highest)**	**1.28**	**1.57 (2nd highest)**
(8-stock portfolios)	**2.01**	**1.67 (Lowest)**	**4.33 (Highest)**	**1.94**	**2.37 (2nd highest)**

Table 5.9. Mean returns, standard deviations and Sharpe ratios under the five different portfolio selection methods for period 5 using period 4's data for estimation of correlation clusters. Portfolios' size was two, four and eight respectively. The risk free rate used to calculate Sharpe ratios was 4.4%.

	2 - stock portfolios	4 – stock portfolios	8 – stock portfolios
AVOVA Test's P value	<0.001	<0.001	<0.001
Levene's Test's P value	<0.001	<0.001	<0.001

The p values of Levene's test and the AVOVA test for period 5.

Period 6	Random	Neighbor-Net Splits graph	Hierarchical Cluster Tree	Minimum Spanning Tree	Industry Groups
Mean return					
(2-stock portfolios)	104.51	115.60	124.81	94.84	107.13
(4-stock portfolios)	104.76	114.77	123.52	95.78	107.09
(8-stock portfolios)	105.77	116.11	122.63	94.40	107.36
Standard Deviation					
(2-stock portfolios)	50.25	44.98	53.80	53.58	52.00
(4-stock portfolios)	33.89	30.02	35.58	34.05	33.83
(8-stock portfolios)	21.33	19.61	20.19	21.90	21.78
Sharpe Ratio					
(2-stock portfolios)	**2.08**	**2.57 (Highest)**	**2.32 (2nd Highest)**	**1.77**	**2.06**
(4-stock portfolios)	**3.04**	**3.77 (Highest)**	**3.42 (2nd Highest)**	**2.76**	**3.12**
(8-stock portfolios)	**4.88**	**5.83 (2nd Highest)**	**5.99 (Highest)**	**4.23**	**4.85**

Table 5.10. Mean returns, standard deviations and Sharpe ratios under the five different portfolio selection methods for period 6 using period 5's data for estimation of correlation clusters. Portfolios' size was two, four and eight respectively. The risk free rate used to calculate Sharpe ratios was 1.7%.

	2 - stock portfolios	4 – stock portfolios	8 – stock portfolios
AVOVA Test's P value	<0.001	<0.001	<0.001
Levene's Test's P value	<0.001	<0.001	<0.001

The p values of Levene's test and the AVOVA test for period 6.

Chapter 6 Extended study of the neighbor-Net method using ASX200

In this chapter we continue to simulate portfolios by picking portfolios randomly and from different correlation clusters using a larger set of data, namely the ASX 200. The "correlation clusters" are defined in four different ways; namely correlation cluster revealed by neighbor-Net splits graph, industry groups, industry groups within correlation clusters, non-industry groups within correlation cluster groups. The detailed illustration of how clusters are defined will be given in the method section. We then discuss and compare the simulation methods.

6.1 Data

6.1.1 Data collection and transformation

We used the stocks' weekly price data in the ASX 200 as our dataset. Weekly prices along with the dividend rate and payment date for the period 03 May 2000 to 04 December 2013 were obtained from DataStream. The names of all stocks and stock ticker symbols plus one or two letters to indicate the industry groups the stocks belong to are listed in Appendix 5. Similar to Chapter 5, we divided the whole period into six shorter periods shown in Figure 6.1 and used also the "out-of-sample" testing method to test the efficiencies of the five methods at diversifying portfolios.

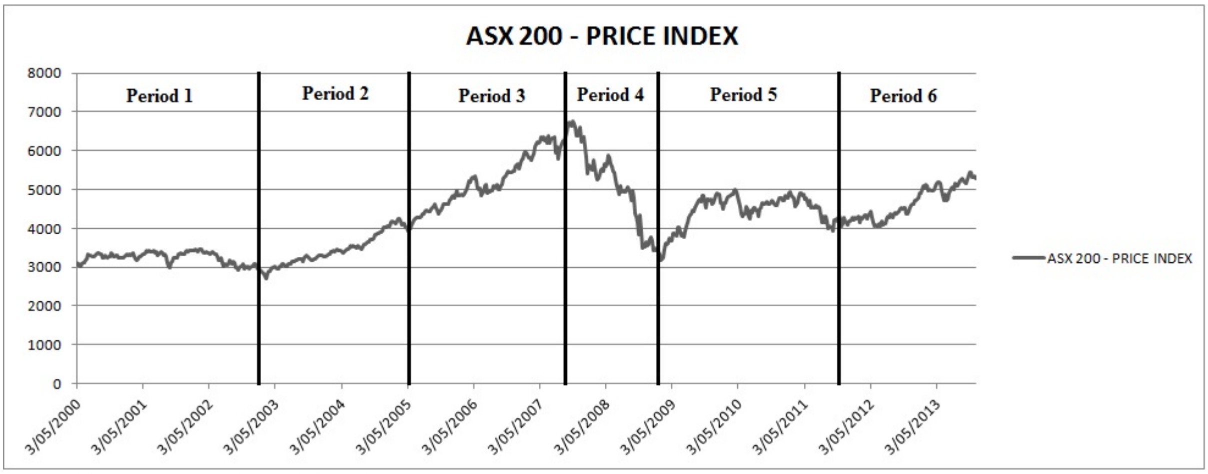


Figure 6.1. The whole period split into 6 shorter periods.

Similar to Chapter 5, we used stocks' weekly return data in period one to determine the clusters, then observed period two's return distributions of the simulated portfolios (1000 replications) which are picked from the different correlation clusters. Because out-of-sample testing was used in our analysis, the simulation then was continued for period three, four, five and six based on the graphs produced from the weekly returns in period two, three, four and five respectively. Therefore, three measurements were needed to carry out the simulations:

1. Individual stock's weekly returns for periods one through five.
2. Individual stock's return for periods two through six.
3. Individual stock's standard deviation of weekly returns for period two through six.

The formulas and methods for observing the above data are explained data section in Chapter 5 therefore are omitted in this section.

6.1.2 Create correlation matrices

We created five correlation matrices (period one through five) using the software package R. (Refer to Section 5.1.2 in Chapter 5 for details).

6.1.3 Transform correlation matrices into distance matrices

Correlations in each correlation matrix were transformed into distance measures using the formula $\sqrt{2(1-\rho_{i,j}})$. After the transformation, the new "correlations" in the matrices ranged from 0 to 2. (Refer to Section 5.1.3 in Chapter 5 for details).

6.2 Method

Since this chapter is an expansion of analysis derived from chapter 5, rationale and motivation behind using simulation method are still the same. We will not describe the method in any details in this chapter. For details concepts behind the simulation method, please refer back to Section 5.2.1 in Chapter 5.

6.2.1 Determine correlation clusters

6.2.1.1 Define correlation clusters and run simulations

Similar to Chapter 5, we used R to simulate 1000 portfolios of two, four and eight stocks each of which was formed by picking stocks:

1. randomly
2. from the correlation clusters that are determined by previous period's neighbor-Net splits graph;

3. from different industry groups; and
4. from the correlation clusters that are defined by both industry groups and the neighbor-Net splits tree
5. from the correlation clusters that are defined by only the neighbor-Net splits graphs but not industry groups

We explain each strategy of picking stocks below:

Strategy # 1: Pick stocks randomly

This method is exactly the same as the method described in Section 5.2.1.3 in Chapter 5 which randomly picks stocks using a uniform distribution without replacement. For details, please refer back to previous chapter.

Strategy # 2: Pick stocks from the correlation clusters that are determined by previous period's neighbor-Net splits graph

Because we use out-of-sample testing to test the efficiency of the visualization method at splitting clusters, we used the neighbor-Net splits graph that was made from previous period's data. That is, use period one, two, three, four and five's data to produce neighbor-Net splits graphs, and test them by examining the period two, three, four, five and six's portfolios returns and standard deviations. The clusters were determined by the structure of neighbor-Net splits graphs which were easily observed from looking at these graphs. (See Figure 6.3 in this chapter and Figure 6.5, 6.7, 6.9 and 6.11 in Appendix 7) To make it comparable to industry selection, we split all graphs into 10 or 11 clusters in every period since there are total 11 industry groups in the whole study period. For portfolios of both four and eight stocks, each stock was picked from different clusters without replacement.

Strategy # 3: Pick stocks from different industry groups

There were 11 industry groups. For portfolios of two stocks, firstly randomly select two industry groups without replacement, then randomly select one stock from each of the selected industry. For portfolios of four stocks, firstly randomly select four industry groups without replacement, then randomly select one stock from each of the selected industry group. Similarly, for portfolios of eight stocks, firstly select eight industry groups without replacement, then randomly select one stock from each of the selected industry group.

Strategy # 4 and # 5 from the correlation clusters that are defined by both the industry groups and the neighbor-Net splits tree; from the correlation clusters that are defined by only the neighbor-Net splits graphs but not industry groups

We discuss strategy #4 and #5 at the same time, because these two strategies are complementary. In particular, we defined clusters using both correlation clusters shown by the neighbor-Net splits graphs as well as one or two industry groups. For instance, in Figure 6.2 there is a correlation cluster group in red. Within this correlation cluster there are four stocks that are also from the same industry group, so these four stocks (RHC_H, CSL_H, COH_H and RMD_H) are defined as one cluster under strategy #4. For strategy #5, using the same correlation cluster in Figure 6.3, the cluster in red contains eight stocks that are in the same cluster but not same industry group (SUN_F, CBA_F, NAB_F, ANZ_F, WBC_F, GNG_CG, BOQ_F and STO_O). Therefore this group fits the group defined in strategy #5; we call this cluster "correlation and non-industry" cluster. It is worth noting in the correlation and non-industry cluster, there are in fact six stocks from the finance industry group including the "four pillars" banks (SUN_F, CBA_F, NAB_F, ANZ_F, WBC_F, BOQ_F) which one may argue they belong to a correlation and industry group. However, since the finance industry is a relatively large industry group, that is, a large number of stocks in our data set are from the finance industry group, we compromise the fact that the correlation cluster and non-industry cluster contains a lot of stocks from the finance industry. In fact, we decided to define the purple finance and technology (with a hyphen of TC) stocks as the correlation and finance (and technology) cluster which contains 12 finance stocks.

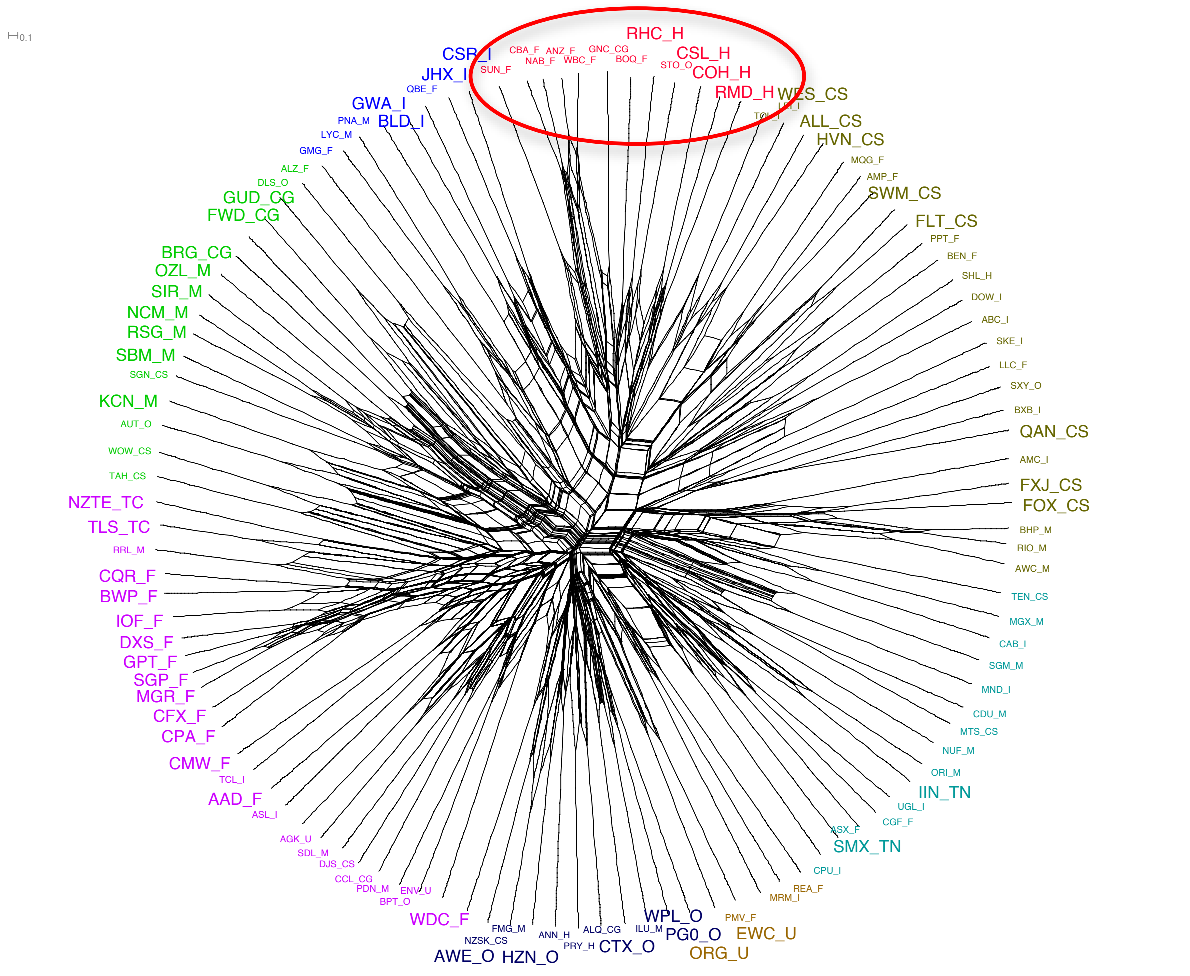


Figure 6.2 Graph demonstrate picking strategy 4 and 5. In the correlation cluster in red, the four stocks: RHC_H, CSL_H, COH_H and RMD_H are in a cluster defined in strategy #4 which is restricted by both correlation cluster and industry group. The complementary stocks: SUN_F, CBA_F, NAB_F, ANZ_F, WBC_F, GNG_CG, BOQ_F and STO_O are in a cluster defined in strategy #5 which is restricted by correlation cluster but not industry cluster.

Therefore for each period, we will end up with five return to density curves and five returns to weekly volatility distribution graphs based on the five stock picking methods. We then discuss and compare them.

6.2.1.2 Determine clusters for periods one to five

The clusters defined by strategies two, four and five for period one to five are defined as follows (strategy #3 is picking stocks from industry groups, industry groups are simple and clear and does not need graphic illustration):

Period 1: 3 May 2000 to 26 March 2003 clusters defined by the neighbor-Net splits graph

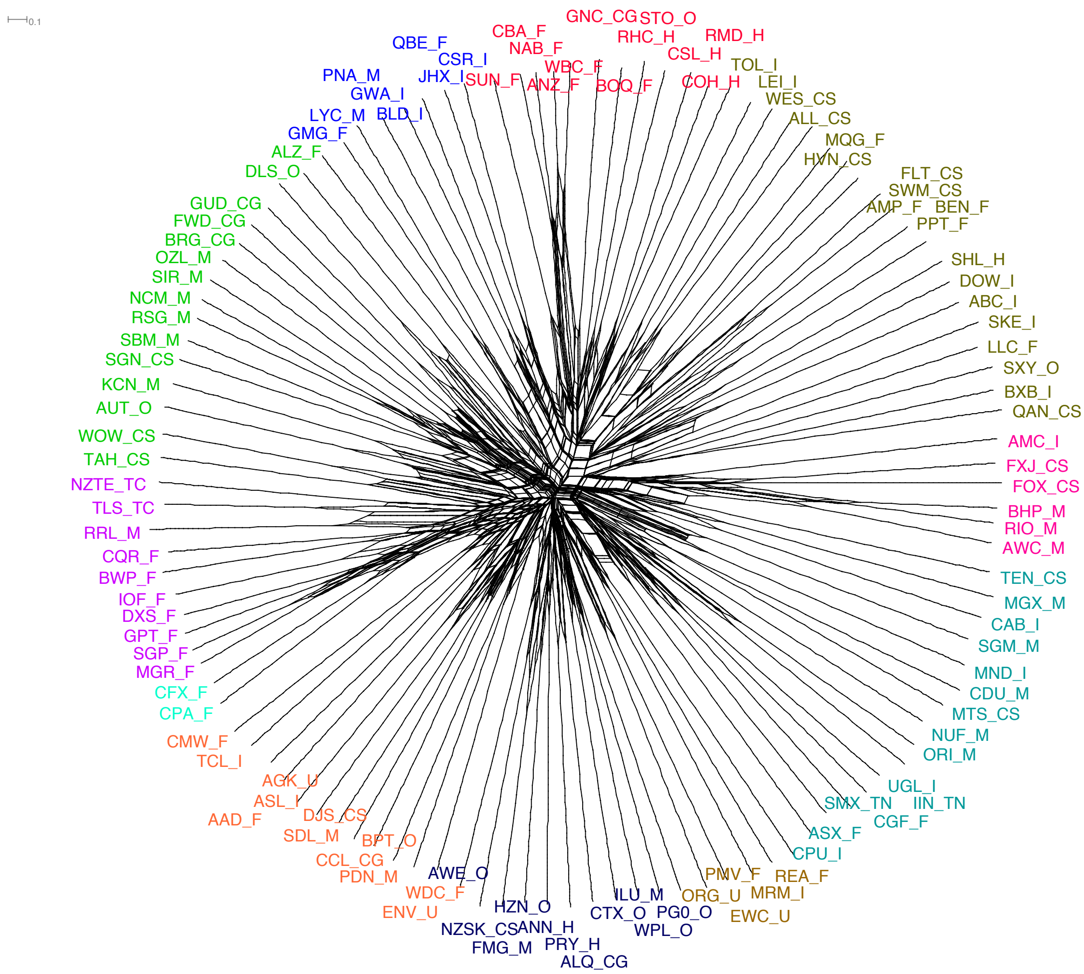


Figure 6.3. The neighbor-Net splits graph split the 115 stocks into 11 correlation clusters defined by strategy #3 for the period 1: 3 May 2000 to 26 March 2003. Some of the clusters were made small to fit with the fact that some of the industry groups are also small.

Correlation Cluster Group	Stocks in each correlation cluster group
Cluster #1	WDC_F, ENV_U,BPT_O, PDN_M(removed as outlier), CCL_CG, DJS_CS, SDL_M, AGK_U, ASL_I, AAD_F, TCL_I, CMW_F
Cluster #2	CFX_F, CPA_F
Cluster #3	MGR_F, SGP_F, GPT_F, DXS_F, IOF_F, BWP_F, CQR_F, RRL_M, TLS_TC, NZTE_TC
Cluster #4	TAH_CS, WOW_CS, AUT_O, KCN_M, SGN_CS, SBM_M, RSG_M, NCM_M, SIR_M, OZL_M, BRG_CG, FWD_CG, GUD_CG, DLS_O, ALZ_F,
Cluster #5	GMG_F, LYC_M, PNA_M, GWA_I, BLD_I, QBE_F, JHX_I, CSR_I
Cluster #6	SUN_F, CBA_F, NAB_F, ANZ_F, WBC_F, GNC_CG, BOQ_F, RHC_H, STO_O, CSL_H, COH_H, RMD_H
Cluster #7	TOL_I, LEI_I, WES_CS, ALL_CS, HVN_CS, MQG_F, AMP_F, SWM_CS, FLT_CS, PPT_F, BEN_F, SHL_H, DOW_I, ABC_I, SKE_I, LLC_F, SXY_O, BXB_I, QAN_CS
Cluster #8	AMC_I, FXJ_CS, FOX_CS, BHP_M, RIO_M, AWC_M
Cluster #9	TEN_CS, MGX_X, CAB_I, SGM_M, MND_I, CDU_M, MTS_CS, NUF_M, ORI_M, IIN_TN, UGL_I, CGF_F, SMX_TN, ASX_F, CPU_I
Cluster #10	REA_F, MRM_I, EWC_U, PMV_F, ORG_U
Cluster #11	PG0_O, WPL_O, ILU_M, CTX_O, ALQ_CG, PRY_H, ANN_H, HZN_O, FMG_M(removed as outlier), NZSK_CS, AWE_O

Table 6.1. The names of stocks in each correlation cluster determined by the neighbor-Net of period 1: 3 May 2000 to 26 March 2003.

Industry Group	**Stocks in each industry group**
Industry Group #1	ALL_CS, FOX_CS, DJS_CS, FXJ_CS, MTS_CS, HVN_CS, FLT_CS, QAN_CS, SGN_CS, SWM_CS, NZSK_CS, TAH_CS, TEN_CS, WES_CS,WOW_CS
Industry Group #2	AGK_U, EWC_U, ENV_U, ORG_U
Industry Group #3	AMP_F, ANZ_F, ASX_F, AAD_F, BWP_F, BOQ_F, BEN_F, ALZ_F, CFX_F, CGF_F, CQR_F, CBA_F, CPA_F, CMW_F, GMG_F, GPT_F, IOF_F, LLC_F, MGR_F, DXS_F, MQG_F, NAB_F, PPT_F, PMV_F, QBE_F, REA_F, SGP_F, SUN_F, WDC_F, WBC_F
Industry Group #4	ALQ_CG, BRG_CG, CCL_CG
Industry Group #5	AWE_O, AUT_O, BPT_O, PG0_O, STO_O, CTX_O, SXY_O, WPL_O, DLS_O, HZN_O
Industry Group #6	ABC_I, AMC_I, ASL_I, BLD_I,BXB_I, CSR_I, CAB_I, GWA_I, CPU_I, DOW_I, JHX_I, SKE_I, TOL_I, TCL_I, UGL_I, LEI_I, MRM_I, MND_I
Industry Group #7	AWC_M, BHP_M, CDU_M, ILU_M, KCN_M, LYC_M, MGX_M, NCM_M, NUF_M, OZL_M, ORI_M, PNA_M, RRL_M, RSG_M, RIO_M, SGM_M, SIR_M, SBM_M, SDL_M
Industry Group #8	ANN_H, CSL_H, COH_H, PRY_H, RHC_H, RMD_H, SHL_H
Industry Group #9	FWD_CG, GUD_CG, GNC_CG
Industry Group #10	NZTE_TC, TLS_TC
Industry Group #11	SMX_TN, IIN_TN

Table 6.2. The code of each stock in the 11 industry groups. The industry groups are consistent over all periods.

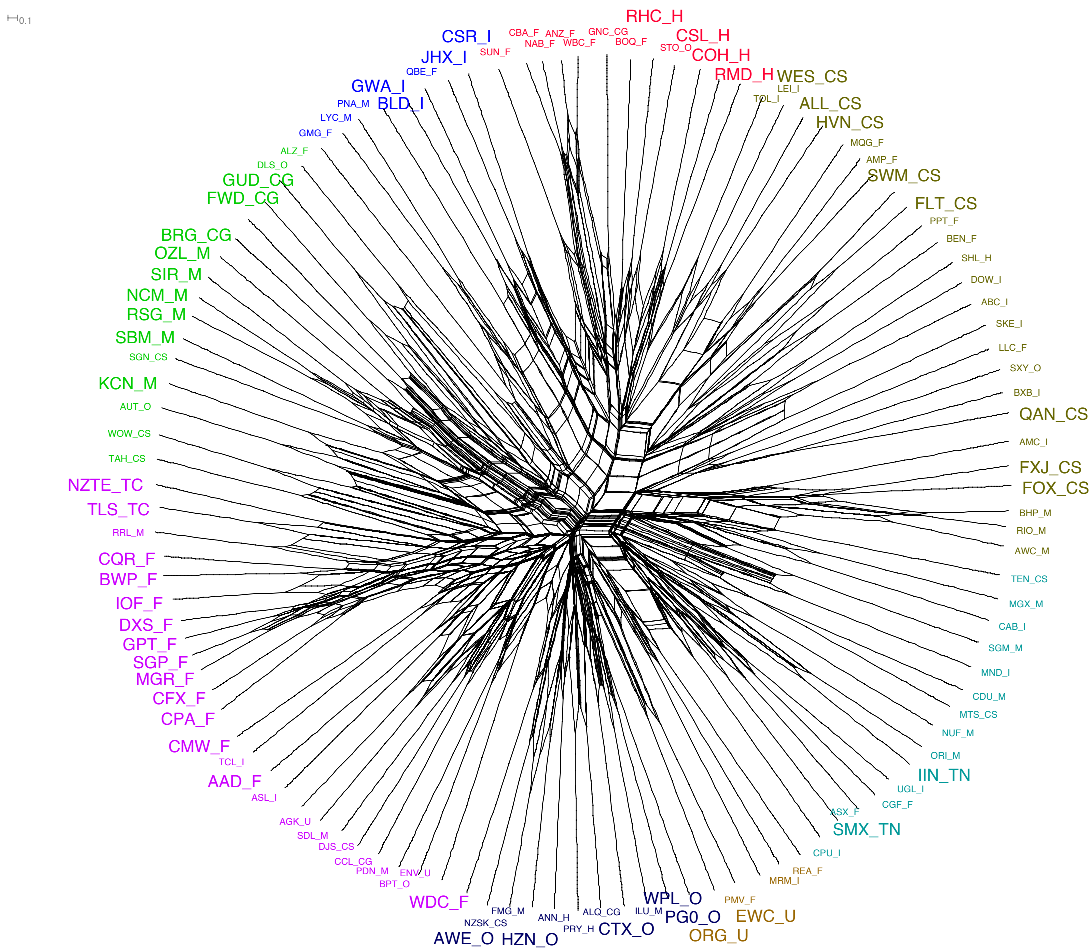


Figure 6.4. The clusters defined by both correlation clusters and industry groups (in bigger size) defined in strategy #4 and the correlation clusters and non-industry groups (in small size) defined in strategy #5 for period 1: 3 May 2000 to 26 March 2003.

Industry Group within correlation cluster group	Names of stocks in each cluster
Cluster #1	WDC_F, AAD_F, CMW_F, CPA_F, CFX_F, MGR_F, SGP_F, GPT_F, DXS_F, IOF_F, BWP_F, CQR_F, TLS_TC, NZTE_TC
Cluster #2	KCN_M, SBM_M, RSG_M, NCM_M, SIR_M, OZL_M, BRG_CG, FWD_CG, GUD_CG
Cluster #3	GWA_I, BLD_I,JHX_I,CSR_I
Cluster #4	RHC_H,CSL_H,COH_H,RMD_H
Cluster #5	WES_CS,ALL_CS, HVN_CS,SWM_CS,FLT_CS, QAN_CS, FXJ_CS, FOX_CS
Cluster #6	IIN_TN, SMX_TN
Cluster #7	EWC_U, ORG_U
Cluster #8	PG0_O, WPL_O, CTX_O, HZN _O, AWE_O

Table 6.3. The code of each stock within each industry plus correlation group.

Correlation without industry cluster group	Names of stocks in each cluster
Cluster #1	ENV_U, BPT_O, CCL_CG, DJS_CS, SDL_M, AGK_U, ASL_I, TCL_I, RRL_M
Cluster #2	TAH_CS, WOW_CS, AUT_O, SGN_CS, DLS_O, ALZ_F
Cluster #3	GMG_F, LYC_M, PNA_M, QBE_F
Cluster#4	SUN_F, CBA_F, NAB_F, ANZ_F, WBC_F, GNC_CG, BOQ_F, STO_O
Cluster #5	TOL_I,LEI_I, MQG_F, AMP_F, PPT_F, BEN_F, SHL_H, DOW_I, ABC_I, SKE_I, LLC_F, SXY_O, BXB_I, AMC_I, BHP_M, RIO_M, AWC_M
Cluster #6	TEN_CS, MGX_M, CAB_I, SGM_M, MND_I, CDU_M, MTS_CS, NUF_M, ORI_M, UGL_I, CGF_F, ASX_F, CPU_I
Cluster #7	REA_F, MRM_I, PMV_F
Cluster #8	ILU_M, ALQ_CG, PRY_H, ANN_H, NZSK_CS

Table 6.4. The code of each stock within non-industry plus correlation cluster group.

6.3 Results and discussion

As can be seen from Tables 6.5 to 6.9, the strategy #4 outperformed the other methods in three of the total five periods, namely period three, five and six. In period 2, the strategy # 5 outperformed the other methods. Look closely, we found in Figure 6.13 (d) in Appendix 7, the return distribution and density curve show that the portfolios picked from strategy #4 has two modes. This dramatically enlarged the standard deviations therefore reduced the Sharpe ratios. This, along with the ANOVA test result, confirmed that we should not over-interpret the Sharpe ratio since there is no reason to expect that the mean returns of each cluster should be different. The period 4 is a period of economic down turn, the industry group selection shown to be a better selection method for this period. Unlike in the Dow Jones Industrial Average studies in Chapter 5, the neighbor-Net method did not always perform better than random or industry group selection methods.

Period 2 Simulation results	Random	Neighbor-Net's correlation cluster	Industry Group	Correlation cluster with industry group	Correlation cluster without industry group
Mean return					
(2-stock portfolios)	104.82	101.07	95.08	105.86	100.10
(4-stock portfolios)	100.80	97.98	91.15	106.85	99.48
(8-stock portfolios)	99.34	99.87	92.06	111.72	97.40
Standard Deviation					
(2-stock portfolios)	82.72	72.05	68.98	86.02	53.39
(4-stock portfolios)	55.37	48.00	44.52	56.77	34.44
(8-stock portfolios)	38.18	35.40	31.38	42.33	22.43
Sharpe Ratio					
(2-stock portfolios)	1.26	1.38	1.36	1.21	**1.85 (Highest)**
(4-stock portfolios)	1.79	2.01	2.01	1.86	**2.84 (Highest)**
(8-stock portfolios)	2.56	2.78	2.89	2.60	**4.28 (Highest)**

Table 6.5 (a). Mean returns, standard deviations and Sharpe ratios under the five different portfolio selection methods for period 2 using period 1's data for estimation of correlation clusters. Portfolios' size was two, four and eight respectively. The risk free rate used to calculate Sharpe ratios was 1.5%.

	2 - stock portfolios	4 – stock portfolios	8 – stock portfolios
AVOVA Test's P value	0.0016	<0.001	<0.001
Levene's Test's P value	<0.001	<0.001	<0.001

Table 6.5 (b). The p values of Levene's test and the AVOVA test for period 2: 2 April 2003 to 04 May 2005.

Period 3 Simulation results	Random	Neighbor-Net's correlation cluster	Industry Group	Correlation cluster with industry group	Correlation cluster without industry group
Mean return					
(2-stock portfolios)	149.76	155.82	132.66	129.31	165.65
(4-stock portfolios)	149.26	157.64	138.29	131.66	160.79
(8-stock portfolios)	147.83	156.63	138.36	132.53	159.41
Standard Deviation					
(2-stock portfolios)	103.12	111.04	105.01	70.77	127.87
(4-stock portfolios)	74.83	79.08	69.10	49.97	82.95
(8-stock portfolios)	52.49	53.37	47.08	35.28	60.81
Sharpe Ratio					
(2-stock portfolios)	1.42	1.37	1.23	**1.78 (Highest)**	1.27
(4-stock portfolios)	1.95	1.95	1.95	**2.57 (Highest)**	1.90
(8-stock portfolios)	2.75	2.87	2.87	**3.66 (Highest)**	2.57

Table 6.6 (a). Mean returns, standard deviations and Sharpe ratios under the five different portfolio selection methods for period 3 using period 2's data for estimation of correlation clusters. Portfolios' size was two, four and eight respectively. The risk free rate used to calculate Sharpe ratios was 3.3%.

	2 - stock portfolios	4 – stock portfolios	8 – stock portfolios
AVOVA Test's P value	<0.001	<0.001	<0.001
Levene's Test's P value	<0.001	<0.001	<0.001

Table 6.6 (b). The p values of Levene's test and the AVOVA test for period 3: 11 May 2005 to 10 October 2007.

Period 4 Simulation results	Random	Neighbor-Net's correlation cluster	Industry Group	Correlation cluster with industry group	Correlation cluster without industry group
Mean return					
(2-stock portfolios)	-47.27	-49.88	-40.96	-49.52	-50.80
(4-stock portfolios)	-46.71	-49.11	-40.70	-49.53	-50.93
(8-stock portfolios)	-47.70	-48.64	-40.64	-49.34	-51.05
Standard Deviation					
(2-stock portfolios)	21.87	21.23	23.40	16.59	18.06
(4-stock portfolios)	15.16	14.86	16.39	11.61	12.32
(8-stock portfolios)	10.20	10.28	10.51	7.34	8.36
Sharpe Ratio					
(2-stock portfolios)	-2.34	-2.54	**-1.92 (Highest)**	-3.23	-3.03
(4-stock portfolios)	-3.34	-3.57	**-2.73 (Highest)**	-4.61	-4.46
(8-stock portfolios)	-5.07	-5.12	**-4.25 (Highest)**	-7.27	-6.58

Table 6.7 (a). Mean returns, standard deviations and Sharpe ratios under the five different portfolio selection methods for period 4 using period 3's data for estimation of correlation clusters. Portfolios' size was two, four and eight respectively. The risk free rate used to calculate Sharpe ratios was 4%.

	2 - stock portfolios	4 – stock portfolios	8 – stock portfolios
AVOVA Test's P value	<0.001	<0.001	<0.001
Levene's Test's P value	<0.001	<0.001	<0.001

Table 6.7 (b). The p values of Levene's test and the AVOVA test for period 4: 17 October 2007 to 4 March 2009.

Period 5 Simulation results	Random	Neighbor-Net's correlation cluster	Industry Group	Correlation cluster with industry group	Correlation cluster without industry group
Mean return					
(2-stock portfolios)	166.09	181.97	171.50	155.57	141.13
(4-stock portfolios)	161.08	173.64	170.34	156.61	143.16
(8-stock portfolios)	162.11	154.07	172.79	156.01	144.89
Standard Deviation					
(2-stock portfolios)	154.04	165.03	146.31	130.31	122.96
(4-stock portfolios)	102.87	113.79	106.21	90.27	86.99
(8-stock portfolios)	74.93	64.04	70.41	61.49	59.20
Sharpe Ratio					
(2-stock portfolios)	1.07	1.10	1.17	**1.19 (Highest)**	1.14
(4-stock portfolios)	1.56	1.52	1.60	**1.73 (Highest)**	1.64
(8-stock portfolios)	2.15	2.39	2.44	**2.52 (Highest)**	2.43

Table 6.8 (a). Mean returns, standard deviations and Sharpe ratios under the five different portfolio selection methods for period 5 using period 4's data for estimation of correlation clusters. Portfolios' size was two, four and eight respectively. The risk free rate used to calculate Sharpe ratios was 0.9%.

	2 - stock portfolios	4 – stock portfolios	8 – stock portfolios
AVOVA Test's P value	<0.001	<0.001	<0.001
Levene's Test's P value	<0.001	<0.001	<0.001

Table 6.8 (b). The p values of Levene's test and the AVOVA test for period 5: 11 March 2009 to 5 October 2011.

Period 6 Simulation results	Random	Neighbor-Net's correlation cluster	Industry Group	Correlation cluster with industry group	Correlation cluster without industry group
Mean return					
(2-stock portfolios)	44.73	51.82	63.73	41.36	51.69
(4-stock portfolios)	47.29	49.75	60.20	40.31	30.81
(8-stock portfolios)	48.24	49.63	63.26	41.30	30.66
Standard Deviation					
(2-stock portfolios)	51.55	52.23	55.55	34.71	43.59
(4-stock portfolios)	36.41	36.41	35.98	23.76	19.91
(8-stock portfolios)	25.12	24.81	24.69	14.90	13.46
Sharpe Ratio					
(2-stock portfolios)	0.86	0.99	1.14	**1.19 (Highest)**	1.18
(4-stock portfolios)	1.29	1.36	1.67	**1.69 (Highest)**	1.54
(8-stock portfolios)	1.91	1.99	2.55	**2.76 (Highest)**	2.26

Table 6.9 (a). Mean returns, standard deviations and Sharpe ratios under the five different portfolio selection methods for period 6 using period 5's data for estimation of correlation clusters. Portfolios' size was two, four and eight respectively. The risk free rate used to calculate Sharpe ratios was 0.2%.

	2 - stock portfolios	4 – stock portfolios	8 – stock portfolios
AVOVA Test's P value	<0.001	<0.001	<0.001
Levene's Test's P value	<0.001	<0.001	<0.001

Table 6.9 (b). The p values of Levene's test and the AVOVA test for period 6: 12 October 2011 to 4 December 2013.

Over all, by combining the neighbor-Net and the knowledge of industry groups' information, we do see risk reductions of portfolios in four out of five periods studied which are period three, four, five and six. It is prudent to respond to our research question that visualization in conjunction of industry group knowledge does give a promising alternative solution to portfolio diversification.

Future research could focus on different data sets and a larger number of periods to examine the neighbor-Net method in conjunction with other types of market data to search further solutions to portfolio diversification.

Chapter 7 A glance at the neighbor-Net splits graphs produced from partial correlation matrices

In previous chapters we examined the clusters revealed by visualizing a correlation matrix. In this chapter, we show that a visualization approach can also help us to reveal whether a different stock eventually controls the observed relationship between any two stocks by studying the statistical measure of partial correlation. In recent years, instead of using full correlation matrices researchers have been using partial correlations to analyse potential structures of stock markets. For example, Zhang et al. (2010) used partial correlation networks to study the influence of the main index on individual stocks. Kenett et al. (2010) used partial correlation planar graphs to study the dominating stocks in financial markets. These papers used MST and/or HCT as a base method for building their networks, and we believe a neighbor-Net splits graph will provide an alternative visual platform to gain a deeper understanding the underlying structure of the financial markets.

A partial correlation coefficient measures the correlation between two variables, when conditioned on one or more other variables. For example, the partial correlation coefficient $\rho(X, Y|I_{ASX200})$ is the correlation between variable X and Y given the ASX200 Index which is the Pearson correlation coefficient between the residuals of X and Y when both are regressed on the ASX200 Index. It removes the effect of ASX200 index on both X and Y and estimates the pure correlation between X and Y without interaction of the ASX200 Index (Equation 1). In financial terms, regressing on ASX200 ought to remove the systematic risk, and in theory everything remaining should be idiosyncratic risk.

Similarly, the partial correlation coefficient $\rho(X, Y|I_{Finance})$ is the correlation between variable X and Y given the Finance Index which is the Pearson correlation coefficient between the residuals of X and Y when are regressed on the Finance Index (Equation 2). In financial terms, regressing on the ASX 200 financials removes the risk factor associated with the financial industry.

$$\rho(X,Y|I_{ASX200}) = \frac{\rho(X,Y)-\rho(X,I_{ASX200})\rho(Y,I_{ASX200})}{\sqrt{[(1-\rho(X,I_{ASX200})^2][(1-\rho(Y,I_{ASX200})^2]}} \tag{1}$$

$$\rho(X,Y|I_{Finance}) = \frac{\rho(X,Y)-\rho(X,I_{Finance})\rho(Y,I_{Finance})}{\sqrt{[(1-\rho(X,I_{Finance})^2][(1-\rho(Y,I_{Finance})^2]}} \tag{2}$$

In this chapter, we briefly explain the neighbor-Net splits graph produced from the partial correlation matrices of period from 3 May 2000 to 26 March 2003 and the period from 11 March 2009 to 05 October 2011. The first period was a pre-crash calm period where the macroeconomic condition was steady growth and no financial crisis was taking place. The second period was post-financial crisis period where the 2008 financial market crash happened a year earlier.

Figure 7.1 (a) and Figure 7.1 (b) indicate partial correlation among all stocks removing the effect of the ASX200 index and the effect of ASX200 finance index respectively. As we can see the structures of the neighbor-Net splits graph in Figure 7.1 (a) is very different to that in Figure 7.1 (b). From Figure 7.1 (a) there are some obvious clusters in the graph. In contrast to Figure 7.1 (a), Figure 7.1 (b), which contains the partial correlation after the effect of finance sector was removed, splits into two distinctive clusters. This may indicate that there are two distinct groups of stocks that reacted to the financial crisis very differently. Looking more closely, we find some stocks show relatively high partial correlation in Figure 7.1 (a) but relatively low partial correlation in Figure 7.1 (b). For example, Stock IOF_F and stock DXS_F showed relatively higher partial correlation in Figure 7.1 (a) than in Figure 7.1 (b).

Figure 7.2 (a) and Figure 7.2 (b) indicate partial correlation among all stocks removing the effect of the ASX200 index and the ASX200 finance index for the period 11 March 2009 to 5 October 2011. Surprisingly, each of the two graphs shows two distinct clusters which may indicate that after the financial crisis; there are two different types of stocks that reacted very differently to the financial market crash in 2008.

The examining of the exact structure of the partial correlations is outside of the scope of this thesis, but adding neighbor-Net splits graphs to the study of partial correlation is definitely beneficial to future research because the ordering behaviour of the neighbor-Net gives a different perspective on how the underlying structure of the stock markets changes over time.

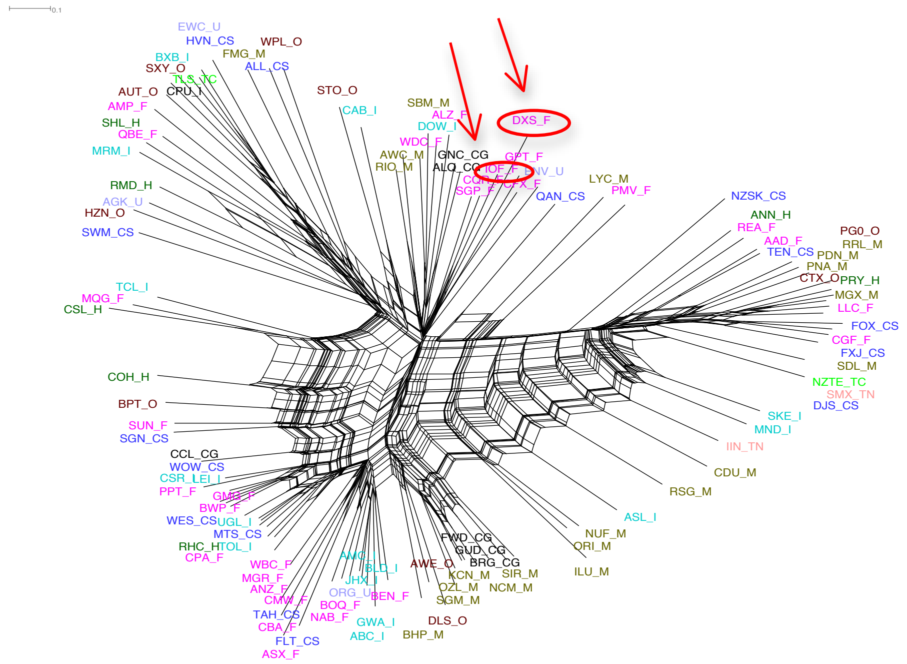


Figure 7.1 (a). Partial correlation of all stocks where the effect of ASX200 Index was removed for the period from 3 May 2000 to 26 March 2003. Different colours represent different industry groups.

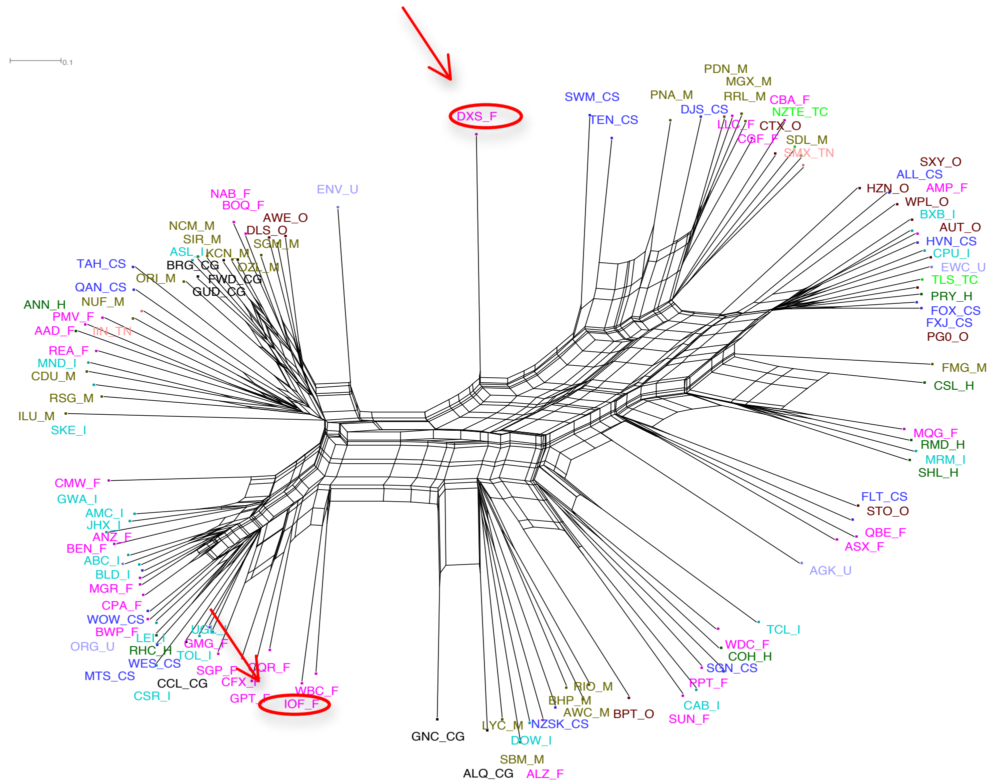


Figure 7.1 (b). Partial correlation among all stocks where the effect of ASX200 Finance Index was removed for the period from 3 May 2000 to 26 March 2003. Different colours represent different industry groups.

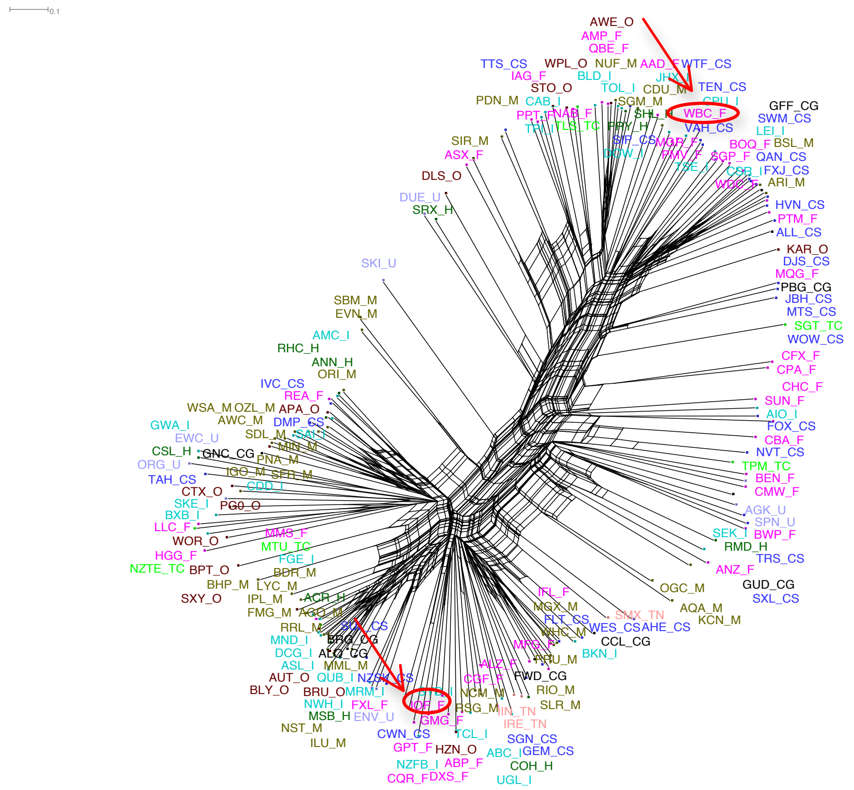


Figure 7.2 (a). Partial correlation among all stocks where the effect of ASX200 Index was removed for the period from 11 March 2009 to 5 October 2011. Different colours represent different industry groups.

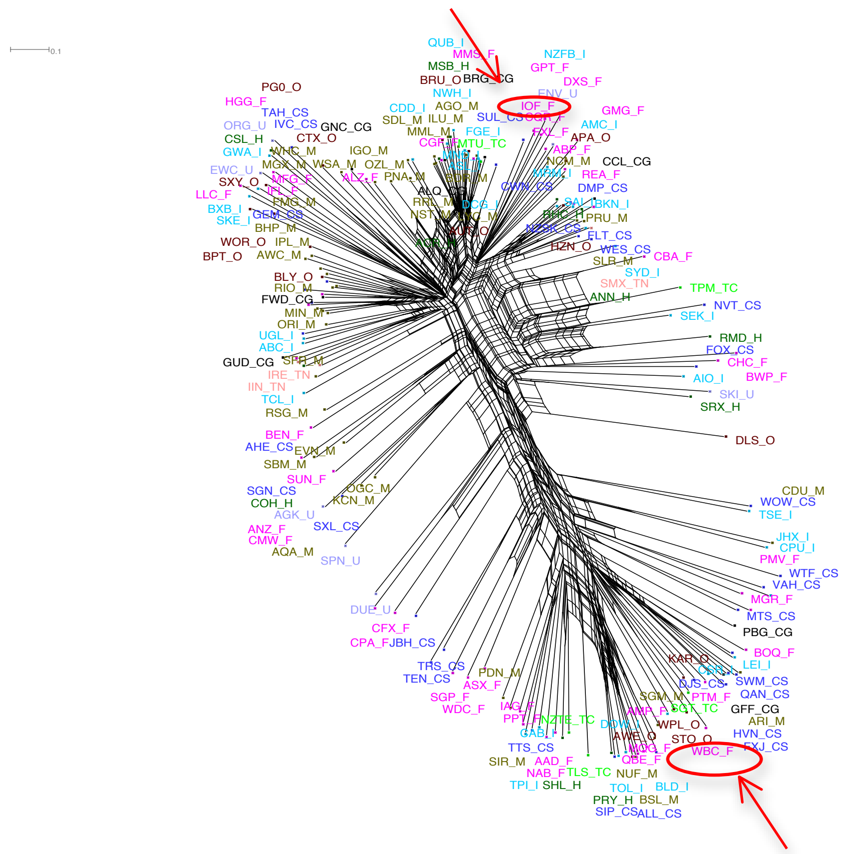


Figure 7.2 (b). Partial correlation among all stocks where the effect of ASX200 Finance Index was removed for the period from 11 March 2009 to 5 October 2011. Different colours represent different industry groups.

Chapter 8 Conclusions and future research

Pragmatically, for portfolio diversification, people define industry groups as correlation cluster groups. That is, to form, for example, a 4-stock portfolio, people pick one stock each from four different industry groups. This is because it was commonly believed that stocks within the same industry groups have similar price movements, while stocks from different industry groups have different price movements (relatively weakly correlated). Industry groups therefore are implicitly defined as correlation cluster groups. In this research, instead of "guessing" what stocks are in the same correlation cluster, we used visualization to attempt actually define the correlation clusters.

8.1 Summary of findings

Two studies were conducted in this research. The first was on the Dow Jones Industrial Average, which contains only 30 stocks, for the period of 20 February 1990 to 14 May 2013. In this set of studies, a total four methods of identifying clusters were studied, plus random selection, therefore a total of five methods of picking portfolios of two, four and eight stocks were studied. The methods compared were picking stocks randomly, picking stocks from different industry groups, picking stocks from different correlation clusters determined by the HCT, picking stocks from different correlation clusters determined by the MST and picking stocks from different correlation clusters determined by the neighbor-Net splits graph. We found that while no single method outperformed the other methods consistently, each method did outperform the other methods in at least one period for at least one portfolio size. This suggests we still have a long way to go in studying portfolio selection methods. We also found the simulated portfolios which were picked from different correlation clusters determined by HCT had the overall highest Sharpe ratios. This phenomenal was due to the fact that the HCT produces highly unbalanced clusters therefore has the potential to produce artificially low standard deviations therefore high Sharpe ratios. The neighbor-Net splits graphs correlation clusters gave the second highest Sharpe ratios for 2-stock sized portfolios in all the periods except for period 5 where the economy suffered a dramatic down turn. Given the fact that the HCT splits highly unbalanced correlation clusters therefore produces artificially low standard deviations, it is appropriate to conclude from this set of studies that neighbor-Net is the best method out of all five methods for determining correlation clusters for portfolio diversification, especially for portfolios of small sizes. It is also for this reason we decided to study the neighbor-Net method in a greater detail in the second study which was on the ASX200.

The second study was conducted on a larger data set - the ASX 200 which contains nearly 200 stocks for the period 03 May 2000 to 04 December 2013. In this study, four methods of splitting stocks into correlation clusters were studied. Taking into account the random selection method, a total of five ways of picking stocks to form portfolios were studied; namely picking stocks randomly, picking stocks from different industry groups, picking stocks from different correlation clusters determined by neighbor-Net splits graphs, picking stocks from different clusters that were bounded by both neighbor-Net splits graphs' correlation cluster groups and industry groups, and finally, picking stocks from clusters that were bounded by neighbor-Net splits graphs' correlation clusters and non-industry groups. The first three methods are comparable because all stocks were eligible to be selected while picking stocks. But they are not comparable with the fourth and fifth method, because the fourth and fifth methods are complimentary; the sum of their available stocks is less or equal to the total number of stocks. We found that the fourth method where clusters were bounded by both neighbor-Net splits graphs' correlation cluster groups and industry groups was definitely the best method among the all five methods. In period three, five and six, the 1000 simulated portfolios picked from the clusters that were bounded by both neighbor-Net splits graphs and industry groups were shown to have the highest Sharpe ratios. In period four, the period of economic down turn, the industry group selection method was shown to be the best method even though the Sharpe ratios were negative. It indicates that we still have a long way to go in diversifying portfolios in a time of economic and stock market downturns where risk reduction is most needed.

In Chapter 7, we briefly looked at two sets of partial correlations which were conditioned on ASX200 index and ASX200 finance index. These graphs of partial correlations contained a large amount of information and future research can look at it in detail, especially the change of structures of stock clusters over time. It could be useful for studying stock market behaviour and individual stock's behaviour.

8.2 Future research

Future work may concentrate on the following areas:

(1) The use of splitstree4 to assign more precise clusters instead of observing by eye should be studied.

(2) The present study did not include expected return as an input to portfolio selection. Adding investors' expectations of returns could improve the return component of the Sharpe ratio and needs further study.

(3) A massive amount of data is generated by splitstree4 while forming the neighbor-Net splits graphs, the usefulness of this data awaits further investigation.

(4) Future research could also concentrate on larger data sets and on a larger number of periods for better comparisons among the selection methods

References

Barber, Brad M., & Odean, Terrance. (2008). All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors. The Review of Financial Studies, 21(2), 785-818.

Bernstein, W. (2001). The Intelligent Asset Allocator. McGraw-Hill.

Bonanno, G., Caldarelli, G., Lillo, F., Micciché, S., Vandewalle, N., & Mantegna, R. N. (2004). Networks of equities in financial markets. The European Physical Journal B, 38(2), 363-371.

Borg, Ingwer, & Groenen, Patrick J. F. (2005). Modern multidimensional scaling: theory and applications. New York: Springer.

Brodie, Joshua, Daubechies, Ingrid, De Mol, Christine, Giannone, Domenico, & Loris, Ignace. (2009). Sparse and stable Markowitz portfolios. Proceedings of the National Academy of Sciences of the United States of America, 106(30), 12267-12272.

Bryant, David, & Moulton, Vincent. (2004). Neighbor-net: an agglomerative method for the construction of phylogenetic networks. Molecular biology and evolution, 21(2), 255-265.

Clarke, Roger, De Silva, Harindra, & Thorley, Steven. (2006). Minimum-variance portfolios in the U.S. equity market. Journal of Portfolio Management, 33(1), 10-24.

Damghani, Babak Mahdavi. (2013). The Non-Misleading Value of Inferred Correlation: An Introduction to the Cointelation Model. Wilmott, 2013(67), 50-61.

Deboeck, Guido J., & Kohonen, Teuvo. (1998). Visual explorations in finance: with self-organizing maps. New York: Springer.

DeMiguel, Victor, Garlappi, Lorenzo, & Uppal, Raman. (2009). Optimal versus Naive Diversification: How Inefficient Is the 1/N Portfolio Strategy? The Review of Financial Studies, 22(5), 1915-1953.

Driessen, Joost, Melenberg, Bertrand, & Nijman, Theo. (2003). Common factors in international bond returns. Journal of International Money and Finance, 22(5), 629-656.

Everitt, Brian. (1978). Graphical techniques for multivariate data. London: Heinemann Educational.

Florian, Wickelmaier. (2003). An Introduction to MDS. Retrieved from homepages.uni-tuebingen.de/florian.wickelmaier/.../Wickelmaier2003SQ.

Huson, Daniel H., & Bryant, David. (2006). Application of phylogenetic networks in evolutionary studies. Molecular biology and evolution, 23(2), 254-267.

Kaufman, Leonard, & Rousseeuw, Peter J. (1990). Finding groups in data: an introduction to cluster analysis. New York: Wiley.

Kenett, Dror Y., Tumminello, Michele, Madi, Asaf, Gur-Gershgoren, Gitit, Mantegna, Rosario N., & Ben-Jacob, Eshel. (2010). Dominating clasp of the financial sector revealed by partial correlation analysis of the stock market. PloS one, 5(12), e15032.

Kruskal, Joseph B., & Wish, Myron. (1978). Multidimensional Scaling (Vol. 11). Thousand Oaks: Sage.

Lee, H. Y., & Ong, H. L. (1996). Visualization support for data mining. IEEE EXPERT-INTELLIGENT SYSTEMS & THEIR APPLICATIONS, 11(5), 69-75.

Lee, Wai. (2011). Risk-Based Asset Allocation: A New Answer to an Old Question? The Journal of Portfolio Management, 37(4), 11-12.

Mantegna, R. N. (1999). Hierarchical structure in financial markets. The European Physical Journal B - Condensed Matter and Complex Systems, 11(1), 193-197.

Markowitz, Harry. (1952). Portfolio selection*. The journal of finance, 7(1), 77-91.

Markowitz, H. (1959). Portfolio selection: efficient diversification of investments. John Wiley & Sons, Inc., New York, Chapman & Hall, Limited, London

Naylor, Michael J., Rose, Lawrence C., & Moyle, Brendan J. (2007). Topology of foreign exchange markets using hierarchical structure methods. Physica A: Statistical Mechanics and its Applications, 382(1), 199-208.

Ong, Hwee-Leng, & Lee, Hing-Yan. (1996). Software report: Winviz—A visual data analysis tool. Computers & Graphics, 20(1), 83-84.

Redpath, Robert. (2000). A Comparative Study of Visualization Techniques for Data Mining. Retrieved from www.csse.monash.edu.au/~srini/theses/Redpath_Thesis.pdf

Rufus G. Rankin, IV. (2003). Improving multiasset portfolio diversification using principle component analysis for investment selection. Retrieved from http://libraryds.grenobleem.com/fr/Publications/Theses%20DBA/Rufus%20G.%20RANKIN.pdf.

Saitou, N., & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular biology and evolution, 4(4), 406-425.

Sharpe, William F. (1966). Mutual Fund Performance. The Journal of Business, 39(1), 119-138.

Sharpe, W. F. (1994). THE SHARPE RATIO. Journal of Portfolio Management, 21(1), 49-58.

Studier, J. A., & Keppler, K. J. (1988). A note on the neighbor-joining algorithm of Saitou and Nei. Molecular biology and evolution, 5(6), 729-731.

The Economist May 3rd 2014 retrieved from http://www.economist.com/news/briefing/21601500-books-and-music-investment-industry-being-squeezed-will-invest-food

Trosset, Michael W. (2005). Visualizing Correlation. Journal of Computational and Graphical Statistics, 14(1), 1-19.

Turner, H. (2011). A peer-reviewed, open-access publication of the R Foundation for Statistical Computing. R JOURNAL, 3(2), 3-3.

Zhang, Jie, Chen, Yejun, & Zhai, Dongsheng. (2010). Network analysis of Shanghai sector in Chinese stock market based on Partial Correlation (Vol. 3, pp. 321-324).

Appendix 1 Dow Jones Industrial Average 30 stocks company names, symbol tickers and industry groups

Company Name	Symbol	Industry Group
UNITED TECHNOLOGIES	UTX	Aerospace
BOEING	BA	Aerospace
ALCOA	AA	Utility
BANK OF AMERICA	BOA	Insurance and Finance
JP MORGAN CHASE & CO.	JPM	Insurance and Finance
WALT DISNEY	DIS	Retailers
WAL MART STORES	WMT	Retailers
CATERPILLAR	CAT	Industrial
E I DU PONT DE NEMOURS	DD	Industrial
HEWLETT-PACKARD	HPQ	Technology
INTEL	INTC	Technology
INTERNATIONAL BUS.MCHS.	IBM	Technology
AMERICAN EXPRESS	AXP	Insurance and Finance
GENERAL ELECTRIC	GE	Retailers
3M	MMM	Retailers
JOHNSON & JOHNSON	JNJ	Healthcare
UNITEDHEALTH GP.	UNH	Healthcare
HOME DEPOT	HD	Retailers
EXXON MOBIL	XOM	Utility
CHEVRON	CVX	Utility
PROCTER & GAMBLE	PG	Retailers
MERCK & CO.	MRK	Healthcare
PFIZER	PFE	Healthcare
TRAVELERS COS.	TRV	Insurance and Finance
MCDONALDS	MCD	Retailers
COCA COLA	CCE	Retailers
MICROSOFT	MSFT	Technology
AT&T	T	Telecom
CISCO SYSTEMS	CRJ	Telecom
VERIZON COMMUNICATIONS	VZWI	Telecom

Appendix 2 Periods returns of Dow Jones Industrial Average 30 stocks in percentage

	AA	XOM	AXP	BOA	IBM	JPM	TRV	CCE	MCD	JNJ	MRK	PFE	UNH	CAT	CVX
Return For Period 2	562	545	537	618	976	573	415	226	114	768	513	987	267	359	461
Return For Period 3	23	27	40	70	-21	41	19	29	-5	2	7	6	64	101	30
Return For Period 4	8	140	61	122	22	88	74	32	91	62	56	4	90	84	150
Return For Period 5	-21	22	-14	-49	61	28	26	63	88	36	38	38	-40	38	65
Return For Period 6	-36	81	109	-14	96	51	155	116	140	111	113	163	127	91	139

	DD	GE	PG	DIS	HD	WMT	MMM	HPQ	INTC	MSFT	UTX	T	CRJ	VZWI	BA
Return For Period 2	286	726	373	54	565	453	421	241	707	1143	617	375	859	473	187
Return For Period 3	34	-1	49	8	-27	-3	72	22	-5	-18	65	-5	20	3	21
Return For Period 4	64	61	62	42	32	1	21	97	-29	24	62	142	2	80	153
Return For Period 5	20	-29	30	-2	7	42	49	36	38	28	47	50	-14	63	-10
Return For Period 6	158	110	102	131	286	97	84	-48	83	56	84	200	5	268	119

Appendix 3 Weekly standard deviations of Dow Jones Industrial Average 30 stocks

	AA	XOM	AXP	BOA	IBM	JPM	TRV	CCE	MCD	JNJ	MRK	PFE	UNH	CAT	CVX
Period 2 Weekly Standard Deviation	0.0466	0.0286	0.0445	0.0448	0.049	0.0472	0.0416	0.0391	0.0394	0.0326	0.0401	0.0434	0.0589	0.0441	0.0303
Period 3 Weekly Standard Deviation	0.0556	0.0372	0.0525	0.0396	0.0522	0.0684	0.0476	0.0346	0.0467	0.0329	0.0442	0.0412	0.0363	0.0433	0.0331
Period 4 Weekly Standard Deviation	0.0342	0.0242	0.0191	0.0171	0.0229	0.0228	0.0263	0.0203	0.0233	0.0165	0.0368	0.03	0.0346	0.0357	0.0254
Period 5 Weekly Standard Deviation	0.078	0.0319	0.0684	0.1198	0.0362	0.0764	0.0415	0.0285	0.0286	0.0233	0.0477	0.0327	0.0625	0.0655	0.0372
Period 6 Weekly Standard Deviation	0.0462	0.0252	0.0359	0.0622	0.0265	0.0403	0.0233	0.0191	0.0194	0.0166	0.0273	0.0263	0.0356	0.0419	0.0298

	DD	GE	PG	DIS	HD	WMT	MMM	HPQ	INTC	MSFT	UTX	T	CRJ	VZWI	BA
Period 2 Weekly Standard Deviation	0.0402	0.0356	0.038	0.045	0.0491	0.0446	0.0366	0.0576	0.0614	0.0529	0.0393	0.0382	0.0704	0.0358	0.0447
Period 3 Weekly Standard Deviation	0.0392	0.0513	0.0255	0.0555	0.0517	0.0349	0.0324	0.0681	0.0715	0.0497	0.0424	0.0588	0.0706	0.0494	0.0428
Period 4 Weekly Standard Deviation	0.0241	0.0193	0.0172	0.0266	0.0254	0.0235	0.0243	0.0347	0.0355	0.0249	0.0224	0.0254	0.0387	0.0232	0.0281
Period 5 Weekly Standard Deviation	0.0496	0.0585	0.0257	0.0434	0.0454	0.027	0.0333	0.043	0.049	0.039	0.0373	0.0317	0.044	0.0334	0.0501
Period 6 Weekly Standard Deviation	0.0331	0.0357	0.0202	0.0297	0.0291	0.0203	0.0255	0.0515	0.034	0.0303	0.029	0.0219	0.0412	0.0226	0.0329

Appendix 4 A correlation matrix of Dow Jones Industrial Average 30 stocks

	AA	XOM	AXP	BOA	IBM	JPM	TRV	CCE	MCD	JNJ	MRK	PFE	UNH	CAT	CVX	DD	GE	PG	DIS	HD	WMT	MMM	HPQ	INTC	MSFT	UTX	T	CRJ	VZWI	BA
AA	1.000	0.121	0.318	0.291	0.324	0.307	0.201	0.222	0.256	0.218	0.209	0.160	0.262	0.446	0.175	0.449	0.475	0.296	0.384	0.282	0.279	0.458	0.268	0.190	0.258	0.451	0.160	0.249	0.136	0.320
XOM	0.121	1.000	0.084	0.174	0.041	0.078	0.202	0.095	0.074	0.112	0.089	0.215	0.094	0.055	0.616	0.298	0.186	0.084	0.016	0.048	0.067	0.136	0.166	0.065	-0.074	0.058	0.153	-0.026	0.232	0.046
AXP	0.318	0.084	1.000	0.436	0.282	0.473	0.403	0.353	0.416	0.294	0.281	0.266	0.378	0.246	0.139	0.290	0.528	0.415	0.370	0.396	0.421	0.320	0.337	0.400	0.393	0.318	0.189	0.341	0.245	0.383
BOA	0.291	0.174	0.436	1.000	0.161	0.579	0.406	0.243	0.342	0.146	0.185	0.176	0.297	0.185	0.133	0.247	0.441	0.242	0.311	0.320	0.309	0.227	0.276	0.329	0.205	0.249	0.187	0.338	0.098	0.294
IBM	0.324	0.041	0.282	0.161	1.000	0.220	0.236	0.093	0.257	0.190	0.287	0.248	0.179	0.211	0.078	0.387	0.309	0.238	0.242	0.207	0.257	0.327	0.315	0.239	0.293	0.272	0.106	0.207	0.121	0.215
JPM	0.307	0.078	0.473	0.579	0.220	1.000	0.326	0.274	0.404	0.182	0.125	0.187	0.241	0.226	0.177	0.251	0.388	0.317	0.379	0.308	0.288	0.208	0.308	0.293	0.328	0.191	0.142	0.292	0.072	0.307
TRV	0.201	0.202	0.403	0.406	0.236	0.326	1.000	0.280	0.374	0.317	0.265	0.260	0.299	0.156	0.129	0.309	0.383	0.339	0.264	0.341	0.335	0.232	0.240	0.298	0.217	0.257	0.353	0.227	0.306	0.247
CCE	0.222	0.095	0.353	0.243	0.093	0.274	0.280	1.000	0.437	0.543	0.539	0.490	0.320	0.165	0.061	0.245	0.428	0.516	0.433	0.500	0.516	0.226	0.299	0.298	0.444	0.267	0.359	0.340	0.335	0.336
MCD	0.256	0.074	0.416	0.342	0.257	0.404	0.374	0.437	1.000	0.277	0.277	0.211	0.245	0.232	0.157	0.309	0.423	0.431	0.506	0.410	0.397	0.315	0.264	0.363	0.378	0.237	0.316	0.240	0.218	0.306
JNJ	0.218	0.112	0.294	0.146	0.190	0.182	0.317	0.543	0.277	1.000	0.700	0.636	0.454	0.173	0.183	0.296	0.412	0.501	0.342	0.502	0.529	0.262	0.367	0.251	0.400	0.292	0.335	0.305	0.312	0.305
MRK	0.209	0.089	0.281	0.185	0.287	0.125	0.265	0.539	0.277	0.700	1.000	0.683	0.321	0.131	0.074	0.316	0.381	0.505	0.270	0.383	0.442	0.233	0.273	0.201	0.311	0.184	0.265	0.212	0.221	0.342
PFE	0.160	0.215	0.266	0.176	0.248	0.187	0.260	0.490	0.211	0.636	0.683	1.000	0.354	0.111	0.174	0.280	0.357	0.445	0.290	0.378	0.484	0.202	0.346	0.194	0.370	0.189	0.227	0.297	0.272	0.294
UNH	0.262	0.094	0.378	0.297	0.179	0.241	0.299	0.320	0.245	0.454	0.321	0.354	1.000	0.178	0.245	0.232	0.379	0.384	0.320	0.477	0.397	0.183	0.241	0.299	0.474	0.233	0.115	0.501	0.183	0.179
CAT	0.446	0.055	0.246	0.185	0.211	0.226	0.156	0.165	0.232	0.173	0.131	0.111	0.178	1.000	0.235	0.310	0.397	0.209	0.346	0.320	0.270	0.282	0.301	0.296	0.262	0.410	0.133	0.250	0.085	0.245
CVX	0.175	0.616	0.139	0.133	0.078	0.177	0.129	0.061	0.157	0.183	0.074	0.174	0.245	0.235	1.000	0.270	0.178	0.096	0.054	0.071	0.104	0.176	0.110	0.182	0.096	0.154	0.200	0.058	0.199	0.030
DD	0.449	0.298	0.290	0.247	0.387	0.251	0.309	0.245	0.309	0.296	0.316	0.280	0.232	0.310	0.270	1.000	0.444	0.343	0.289	0.278	0.292	0.519	0.371	0.308	0.291	0.352	0.301	0.243	0.299	0.314
GE	0.475	0.186	0.528	0.441	0.309	0.388	0.383	0.428	0.423	0.412	0.381	0.357	0.379	0.397	0.178	0.444	1.000	0.422	0.463	0.419	0.440	0.505	0.428	0.399	0.338	0.495	0.319	0.298	0.308	0.484
PG	0.296	0.084	0.415	0.242	0.238	0.317	0.339	0.516	0.431	0.501	0.505	0.445	0.384	0.209	0.096	0.343	0.422	1.000	0.466	0.471	0.615	0.380	0.349	0.294	0.398	0.283	0.307	0.323	0.307	0.492
DIS	0.384	0.016	0.370	0.311	0.242	0.379	0.264	0.433	0.506	0.342	0.270	0.290	0.320	0.346	0.054	0.289	0.463	0.466	1.000	0.377	0.449	0.338	0.315	0.344	0.391	0.296	0.187	0.364	0.204	0.376
HD	0.282	0.048	0.396	0.320	0.207	0.308	0.341	0.500	0.410	0.502	0.383	0.378	0.477	0.320	0.071	0.278	0.419	0.471	0.377	1.000	0.718	0.264	0.471	0.427	0.605	0.249	0.203	0.516	0.194	0.350
WMT	0.279	0.067	0.421	0.309	0.257	0.288	0.335	0.516	0.397	0.529	0.442	0.484	0.397	0.270	0.104	0.292	0.440	0.615	0.449	0.718	1.000	0.290	0.394	0.378	0.508	0.259	0.244	0.395	0.231	0.360
MMM	0.458	0.136	0.320	0.227	0.327	0.208	0.232	0.226	0.315	0.262	0.233	0.202	0.183	0.282	0.176	0.519	0.505	0.380	0.338	0.264	0.290	1.000	0.185	0.193	0.223	0.387	0.278	0.144	0.306	0.264
HPQ	0.268	0.166	0.337	0.276	0.315	0.308	0.240	0.299	0.264	0.367	0.273	0.346	0.241	0.301	0.110	0.371	0.428	0.349	0.315	0.471	0.394	0.185	1.000	0.473	0.404	0.368	0.095	0.377	0.101	0.273
INTC	0.190	0.065	0.400	0.329	0.239	0.293	0.298	0.298	0.363	0.251	0.201	0.194	0.299	0.296	0.182	0.308	0.399	0.294	0.344	0.427	0.378	0.193	0.473	1.000	0.551	0.227	0.189	0.498	0.211	0.273
MSFT	0.258	-0.074	0.393	0.205	0.293	0.328	0.217	0.444	0.378	0.400	0.311	0.370	0.474	0.262	0.096	0.291	0.338	0.398	0.391	0.605	0.508	0.223	0.404	0.551	1.000	0.245	0.180	0.546	0.129	0.370
UTX	0.451	0.058	0.318	0.249	0.272	0.191	0.257	0.267	0.237	0.292	0.184	0.189	0.233	0.410	0.154	0.352	0.495	0.283	0.296	0.249	0.259	0.387	0.368	0.227	0.245	1.000	0.220	0.233	0.146	0.366
T	0.160	0.153	0.189	0.187	0.106	0.142	0.353	0.359	0.316	0.335	0.265	0.227	0.115	0.133	0.200	0.301	0.319	0.307	0.187	0.203	0.244	0.278	0.095	0.189	0.180	0.220	1.000	0.008	0.671	0.306
CRJ	0.249	-0.026	0.341	0.338	0.207	0.292	0.227	0.340	0.240	0.305	0.212	0.297	0.501	0.250	0.058	0.243	0.298	0.323	0.364	0.516	0.395	0.144	0.377	0.498	0.546	0.233	0.008	1.000	0.078	0.247
VZWI	0.136	0.232	0.245	0.098	0.121	0.072	0.306	0.335	0.218	0.312	0.221	0.272	0.183	0.085	0.199	0.299	0.308	0.307	0.204	0.194	0.231	0.306	0.101	0.211	0.129	0.146	0.671	0.078	1.000	0.157
BA	0.320	0.046	0.383	0.294	0.215	0.307	0.247	0.336	0.306	0.305	0.342	0.294	0.179	0.245	0.030	0.314	0.484	0.492	0.376	0.350	0.360	0.264	0.273	0.273	0.370	0.366	0.306	0.247	0.157	1.000

Appendix 5 AXS 200 Company names, Symbol ticker code and industry groups

COMPANY NAME	ISIN CODE*[1]	ICB INDUSTRY NAME
ABACUS PROPERTY GROUP	ABP_F	Financials
ACRUX LIMITED	ACR_H	Health Care
ADELAIDE BRIGHTON LTD.	ABC_I	Industrials
AGL ENERGY LIMITED	AGK_U	Utilities
ALACER GOLD CORP.	AQG_M	Basic Materials
ALS LTD.	ALQ_CG	Consumer Goods
ALUMINA LTD.	AWC_M	Basic Materials
AMCOR LTD.	AMC_I	Industrials
AMP LTD.	AMP_F	Financials
ANSELL LTD.	ANN_H	Health Care
APA GROUP	APA_O	Oil & Gas
AQUILA RESOURCES LIMITED	AQA_M	Basic Materials
ARDENT LEISURE GROUP	AAD_F	Financials
ARISTOCRAT LEISURE LTD.	ALL_CS	Consumer Services
ARRIUM LTD.	ARI_M	Basic Materials
ASCIANO LTD.	AIO_I	Industrials
ASX LIMITED	ASX_F	Financials
ATLAS IRON LIMITED	AGO_M	Basic Materials
AURIZON HOLDINGS LTD.	AZJ_I	Industrials
AURORA OIL & GAS LTD.	AUT_O	Oil & Gas
AUS.AND NZ.BANKING GLD.	ANZ_F	Financials
AUSDRILL LIMITED	ASL_I	Industrials
AUSTRALAND PR.GP.	ALZ_F	Financials
AUTOMOTIVE HDG.GP.LTD.	AHE_CS	Consumer Services
AWE LIMITED	AWE_O	Oil & Gas
BANK OF QUEENSLAND LTD.	BOQ_F	Financials
BEACH ENERGY LIMITED	BPT_O	Oil & Gas
BEADELL RESOURCES LTD.	BDR_M	Basic Materials
BENDIGO & ADEL.BK.LTD.	BEN_F	Financials
BHP BILLITON LIMITED	BHP_M	Basic Materials
BLUESCOPE STEEL LTD.	BSL_M	Basic Materials
BOART LONGYEAR LTD.	BLY_O	Oil & Gas
BORAL LTD.	BLD_I	Industrials
BRADKEN LIMITED	BKN_I	Industrials
BRAMBLES LTD.	BXB_I	Industrials
BREVILLE GROUP LIMITED	BRG_CG	Consumer Goods
BURU ENERGY LIMITED	BRU_O	Oil & Gas
BWP TRUST	BWP_F	Financials
CABCHARGE AUSTRALIA LTD.	CAB_I	Industrials
CALTEX AUSTRALIA LTD.	CTX_O	Oil & Gas

[1] The one or two letters after the hyphen indicate the industry group the stock belongs to.

Appendix 5 Continued.

CARDNO LIMITED	CDD_I	Industrials
CARSALES.COM LIMITED	CRZ_CS	Consumer Services
CFS RETAIL PR.TST.GROUP	CFX_F	Financials
CHALLENGER LTD.	CGF_F	Financials
CHARTER HALL GROUP	CHC_F	Financials
CHARTER HALL RETAIL REIT	CQR_F	Financials
COCA-COLA AMATIL LTD.	CCL_CG	Consumer Goods
COCHLEAR LIMITED	COH_H	Health Care
COMMONWEALTH BK.OF AUS.	CBA_F	Financials
COMMONWEALTH PR.OFFE.FD.	CPA_F	Financials
COMPUTERSHARE LTD.	CPU_I	Industrials
CROMWELL PROPERTY GROUP	CMW_F	Financials
CROWN RESORTS LTD.	CWN_CS	Consumer Services
CSL LTD.	CSL_H	Health Care
CSR LIMITED	CSR_I	Industrials
CUDECO LTD.	CDU_M	Basic Materials
DAVID JONES LTD.	DJS_CS	Consumer Services
DECMIL GROUP LIMITED	DCG_I	Industrials
DEXUS PROPERTY GROUP	DXS_F	Financials
DOMINO'S PZA.ENTS.LTD.	DMP_CS	Consumer Services
DOWNER EDI LIMITED	DOW_I	Industrials
DRILLSEARCH ENERGY LTD.	DLS_O	Oil & Gas
DUET GROUP	DUE_U	Utilities
DULUXGROUP LTD.	DLX_I	Industrials
ECHO ENTERTAINMENT GLD.	EGP_CS	Consumer Services
ENERGY WORLD CORP.LTD.	EWC_U	Utilities
ENVESTRA LIMITED	ENV_U	Utilities
EVOLUTION MINING LTD.	EVN_M	Basic Materials
FAIRFAX MEDIA LIMITED	FXJ_CS	Consumer Services
FEDERATION CENTRES	FDC_F	Financials
FLEETWOOD CORP.LIMITED	FWD_CG	Consumer Goods
FLETCHER BUILDING LTD.	NZFB_I	Industrials
FLEXIGROUP LIMITED	FXL_F	Financials
FLIGHT CTR.TRVL.GP.LTD.	FLT_CS	Consumer Services
FORGE GROUP LIMITED	FGE_I	Industrials
FORTESCUE METALS GP.LTD.	FMG_M	Basic Materials
G8 EDUCATION LTD.	GEM_CS	Consumer Services
GOODMAN FIELDER LTD.	GFF_CG	Consumer Goods
GOODMAN GROUP	GMG_F	Financials
GPT GROUP	GPT_F	Financials
GRAINCORP LIMITED	GNC_CG	Consumer Goods
GUD HOLDINGS LTD.	GUD_CG	Consumer Goods
GWA GROUP LTD.	GWA_I	Industrials
HARVEY NORMAN HDG.LTD.	HVN_CS	Consumer Services
HENDERSON GROUP PLC.	HGG_F	Financials
HORIZON OIL LTD.	HZN_O	Oil & Gas
IINET LTD.	IIN_TN	Technology

Appendix 5 Continued.

ILUKA RESOURCES LTD.	ILU_M	Basic Materials
INCITEC PIVOT LTD.	IPL_M	Basic Materials
INDEPENDENCE GROUP NL	IGO_M	Basic Materials
INSURANCE AUS.GROUP LTD.	IAG_F	Financials
INVESTA OFFICE FUND	IOF_F	Financials
INVOCARE LIMITED	IVC_CS	Consumer Services
IOOF HOLDINGS LIMITED	IFL_F	Financials
IRESS LTD.	IRE_TN	Technology
JAMES HARDIE INDS.PLC.	JHX_I	Industrials
JB HI-FI LIMITED	JBH_CS	Consumer Services
KAROON GAS AUS.LTD.	KAR_O	Oil & Gas
KATHMANDU HOLDINGS LTD.	NZKM_CS	Consumer Services
KINGSGATE CONS.LTD.	KCN_M	Basic Materials
LEIGHTON HOLDINGS LTD.	LEI_I	Industrials
LEND LEASE GROUP	LLC_F	Financials
LYNAS CORPORATION LTD.	LYC_M	Basic Materials
M2 TELECOM.GP.LTD.	MTU_TC	Telecommunications
MACQUARIE GROUP LTD.	MQG_F	Financials
MAGELLAN FINL.GP.LTD.	MFG_F	Financials
MCMILLAN SHAKESPEARE LTD	MMS_F	Financials
MEDUSA MINING LIMITED	MML_M	Basic Materials
MERMAID MARINE AUS.LTD.	MRM_I	Industrials
MESOBLAST LTD.	MSB_H	Health Care
METCASH LTD.	MTS_CS	Consumer Services
MINERAL RESOURCES LTD.	MIN_M	Basic Materials
MIRVAC GROUP	MGR_F	Financials
MONADELPHOUS GROUP LTD.	MND_I	Industrials
MOUNT GIBSON IRON LTD.	MGX_M	Basic Materials
MQR.ATLAS ROADS GROUP	MQA_I	Industrials
MYER HOLDINGS LTD.	MYR_CS	Consumer Services
NATIONAL AUS.BANK LTD.	NAB_F	Financials
NAVITAS LIMITED	NVT_CS	Consumer Services
NEWCREST MINING LTD.	NCM_M	Basic Materials
NEWS CORP.	NWS_CS	Consumer Services
NORTHERN STAR RES.LTD.	NST_M	Basic Materials
NRW HOLDINGS LIMITED	NWH_I	Industrials
NUFARM LIMITED	NUF_M	Basic Materials
OCEANAGOLD CORP.	OGC_M	Basic Materials
OIL SEARCH LTD.	PG0_O	Oil & Gas
ORICA LTD.	ORI_M	Basic Materials
ORIGIN ENERGY LTD.	ORG_U	Utilities
OZ MINERALS LTD.	OZL_M	Basic Materials
PACIFIC BRANDS LTD.	PBG_CG	Consumer Goods
PALADIN ENERGY LTD.	PDN_M	Basic Materials
PANAUST LIMITED	PNA_M	Basic Materials
PERPETUAL LTD.	PPT_F	Financials
PERSEUS MINING LIMITED	PRU_M	Basic Materials

Appendix 5 Continued.

PLATINUM ASSET MAN.LTD.	PTM_F	Financials
PREMIER INVESTMENTS LTD.	PMV_F	Financials
PRIMARY HEALTH CARE LTD.	PRY_H	Health Care
QANTAS AIRWAYS LIMITED	QAN_CS	Consumer Services
QBE INSURANCE GROUP LTD.	QBE_F	Financials
QUBE HOLDINGS LTD.	QUB_I	Industrials
RAMSAY HEALTH CARE LTD.	RHC_H	Health Care
REA GROUP LIMITED	REA_F	Financials
REGIS RESOURCES LIMITED	RRL_M	Basic Materials
RESMED INCO.	RMD_H	Health Care
RESOLUTE MINING LTD.	RSG_M	Basic Materials
RIO TINTO LIMITED	RIO_M	Basic Materials
SAI GLOBAL LIMITED	SAI_I	Industrials
SANDFIRE RESOURCES NL	SFR_M	Basic Materials
SANTOS LTD.	STO_O	Oil & Gas
SEEK LTD.	SEK_I	Industrials
SENEX ENERGY LTD.	SXY_O	Oil & Gas
SEVEN GROUP HDG.LTD.	SVW_CS	Consumer Services
SEVEN WEST MEDIA LIMITED	SWM_CS	Consumer Services
SHOP.CENTS.AUSAN.PR.GP.	SCP_F	Financials
SIGMA PHARMS.LTD.	SIP_CS	Consumer Services
SILVER LAKE RES.LTD.	SLR_M	Basic Materials
SIMS METAL MAN.LTD.	SGM_M	Basic Materials
SINGAPORE TELECOM.LTD.	SGT_TC	Telecommunications
SIRIUS RESOURCES NL	SIR_M	Basic Materials
SIRTEX MEDICAL LIMITED	SRX_H	Health Care
SKILLED GROUP LTD.	SKE_I	Industrials
SKY NETWORK TV.LTD.	NZSK_CS	Consumer Services
SMS MAN.& TECH.LIMITED	SMX_TN	Technology
SONIC HEALTHCARE LIMITED	SHL_H	Health Care
SOUTHERN CROSS MDA.GLD.	SXL_CS	Consumer Services
SP AUSNET	SPN_U	Utilities
SPARK INFRASTRUCTURE GP.	SKI_U	Utilities
ST BARBARA LIMITED	SBM_M	Basic Materials
STOCKLAND	SGP_F	Financials
STW COMMUNICATIONS GLD.	SGN_CS	Consumer Services
SUNCORP GROUP LTD.	SUN_F	Financials
SUNDANCE RESOURCES LTD.	SDL_M	Basic Materials
SUPER RETAIL GROUP LTD.	SUL_CS	Consumer Services
SYDNEY AIRPORT	SYD_I	Industrials
TABCORP HOLDINGS LTD.	TAH_CS	Consumer Services
TATTS GROUP LIMITED	TTS_CS	Consumer Services
TELECOM CORP.OF NZ.LTD.	NZTE_TC	Telecommunications
TELSTRA CORPORATION LTD.	TLS_TC	Telecommunications
TEN NETWORK HDG.LTD.	TEN_CS	Consumer Services
THE REJECT SHOP LIMITED	TRS_CS	Consumer Services
TOLL HOLDINGS LTD.	TOL_I	Industrials

Appendix 5 Continued.

TPG TELECOM LIMITED	TPM_TC	Telecommunications
TRADE ME GROUP LTD.	NZTM_CS	Consumer Services
TRANSFIELD SERVICES LTD.	TSE_I	Industrials
TRANSPACIFIC INDS.GLD.	TPI_I	Industrials
TRANSURBAN GROUP	TCL_I	Industrials
TREASURY WINE ESTS.LTD.	TWE_CG	Consumer Goods
TWENTY-FIRST CENTURY FOX	FOX_CS	Consumer Services
UGL LIMITED	UGL_I	Industrials
VIRGIN AUS.HOLDINGS LTD.	VAH_CS	Consumer Services
WESFARMERS LTD.	WES_CS	Consumer Services
WESTERN AREAS LTD.	WSA_M	Basic Materials
WESTFIELD GP.	WDC_F	Financials
WESTFIELD RETAIL TRUST	WRT_F	Financials
WESTPAC BANKING CORP.	WBC_F	Financials
WHITEHAVEN COAL LIMITED	WHC_M	Basic Materials
WOODSIDE PETROLEUM LTD.	WPL_O	Oil & Gas
WOOLWORTHS LTD.	WOW_CS	Consumer Services
WORLEYPARSONS LTD.	WOR_O	Oil & Gas
WOTIF COM HOLDINGS LTD.	WTF_CS	Consumer Services

Industry Groups	Stocks in each Industry Group
Industry Group #1	GMG_F, CFX_F, GPT_F, WDC_F, CQR_F, DXS_F, MGR_F, SGP_F, CPA_F, IOF_F, REA_F, CMW_F, ABP_F, AAD_F, LLC_F, QBE_F, MQG_F, ALZ_F, SUN_F, PPT_F, IAG_F, ASX_F, CGF_F, AMP_F, CBA_F, NAB_F, ANZ_F, WBC_F, BEN_F, BOQ_F, PMV_F, BWP_F
Industry Group #2	RHC_H, SRX_H, RMD_H, PRY_H, CSL_H, SHL_H, COH_H, ANN_H
Industry Group #3	MND_I, TSE_I, CAB_I, DOW_I, SKE_I, MRM_I, ASL_I, TCL_I, SYD_I, CSR_I, UGL_I, LEI_I, ORI_M, BLD_I, ABC_I, GWA_I, JHX_I, AMC_I, CPU_I, TOL_I, BXB_I, NZFB_I
Industry Group #4	FWD_CG, GUD_CG, BRG_CG, ALQ_CG, GNC_CG, CCL_CG
Industry Group #5	BSL_M, SIR_M, SBM_M, EVN_M, AQA_M, IGO_M, WSA_M, LYC_M, CDU_M, PDN_M, NCM_M, KCN_M, RRL_M, RSG_M, OZL_M, PNA_M, RIO_M, BHP_M, AWC_M, SDL_M, MGX_M, ILU_M, FMG_M, SGM_M, ARI_M, NUF_M
Industry Group #6	WOR_O, APA_O, AUT_O, DLS_O, HZN_O, SXY_O, BPT_O, AWE_O, PG0_O, WPL_O, STO_O, CTX_O
Industry Group #7	IRE_TN, SMX_TN, IIN_TN
Industry Group #8	SIP_CS, NZSK_CS, DJS_CS, FLT_CS, ALL_CS, QAN_CS, WES_CS, TAH_CS, FOX_CS, FXJ_CS, SWM_CS, HVN_CS, WOW_CS, MTS_CS, SGN_CS, TEN_CS
Industry Group #9	TPM_TC, SGT_TC, TLS_TC, NZTE_TC
Industry Group #10	AGK_U, EWC_U, ORG_U, ENV_U

Stocks in each industry group.

Appendix 6 Figures and Tables for Chapter 5

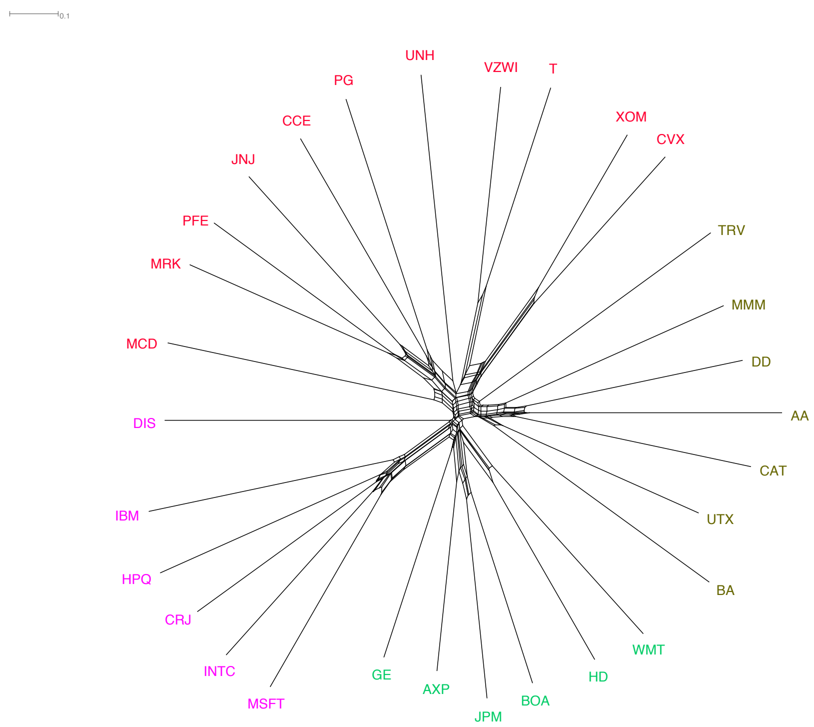


Figure 5.7 (b) Neighbor-Net splits graph produced from stocks' weekly returns in period 2: 11 January 1994 to 01 January 2002.

Cluster 1 (in Red)	MCD, MRK, PFE, JNJ, CCE, PG, UNH, VZWI, T, XOM, CVX
Cluster 2 (in purple)	DIS, IBM,HPQ, CRJ, INTC, MSFT
Cluster 3 (in green)	GE, AXP, JPM, BOA, HD, WMT
Cluster 4 (in dark green)	BA, UTX, CAT, AA, DD, MMM, TRV

Table 5.2 (a). The 4 clusters determined by the neighbor-Net splits graph in period 2: 11 January 1994 to 01 January 2002.

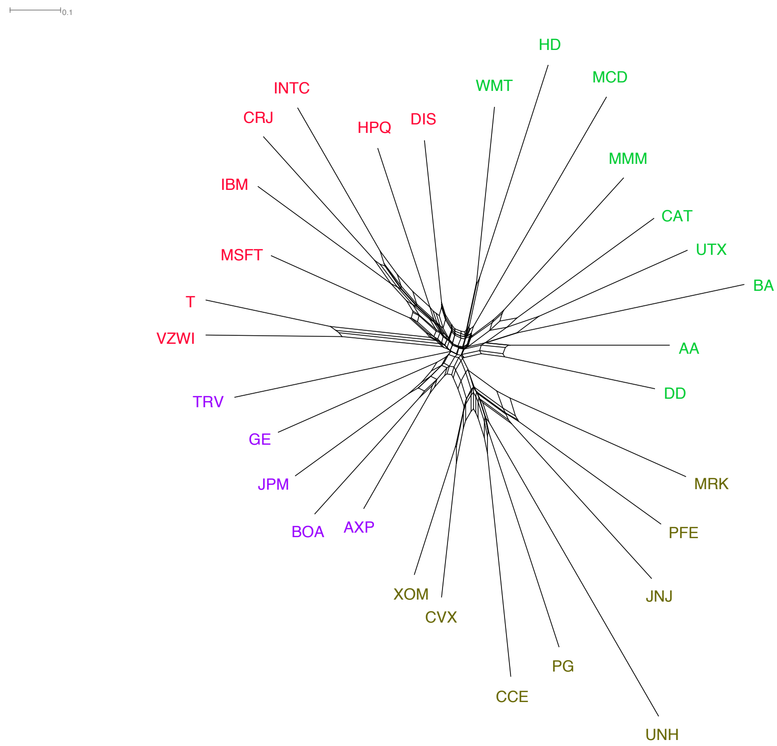


Figure 5.7 (c) Neighbor-Net splits graph produced from stocks' weekly returns in period 3: 08 January 2002 to 06 January 2004.

Cluster 1 (in Red)	VZWI, T, MSFT, IBM, CRJ, INTC, HPQ, DIS
Cluster 2 (in purple)	TRV, GE, JPM, BOA, AXP
Cluster 3 (in dark green)	XOM, CVX, CCE, PG, UNH, JNJ, PFE, MRK
Cluster 4 (in green)	DD, AA, BA, UTX, CAT, MMM, MCD, HD, WMT

Table 5.3 (a). The 4 clusters determined by the neighbor-Net splits graph in period 3: 08 January 2002 to 06 January 2004.

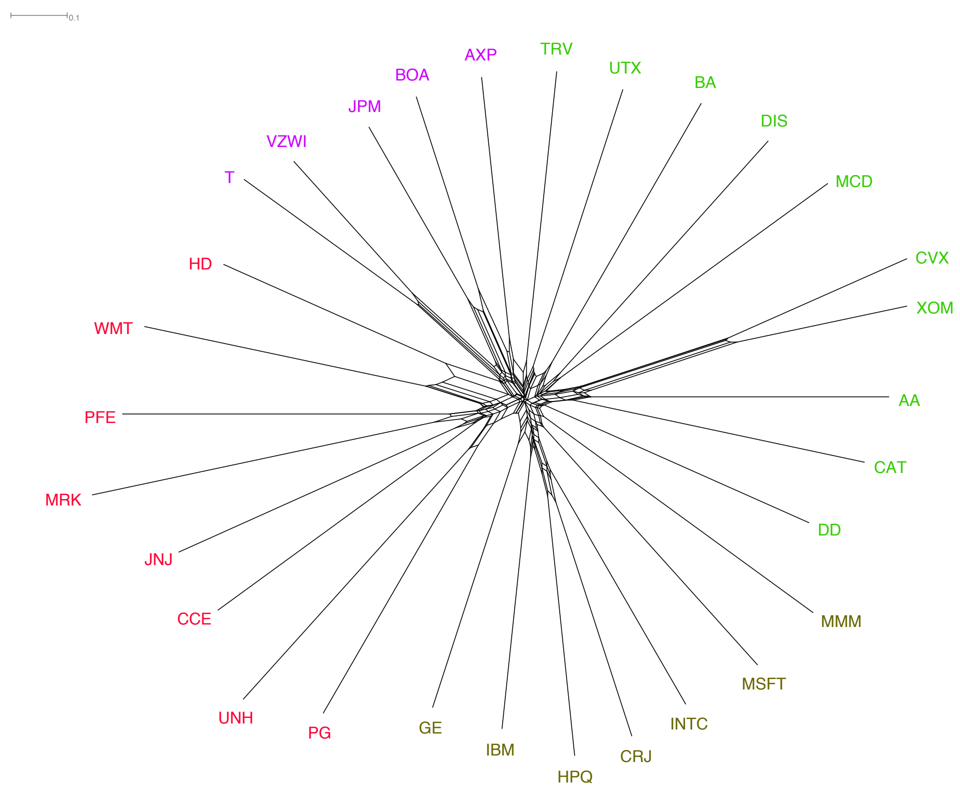


Figure 5.7 (d) Neighbor-Net splits graph produced from stocks' weekly returns in period 4: 13 January 2004 to 02 January 2007.

Cluster 1 (in red)	HD, WMT, PFE, MRK, JNJ, CCE, UNH, PG
Cluster 2 (in dark green)	GE, IBM, HPQ, CRJ, INTC, MSFT, MMM
Cluster 3 (in green)	DD, CAT, AA, XOM, CVX, MCD, DIS, BA, UTX, TRV
Cluster 4 (in purple)	T, VZWI, JPM, BOA, AXP

Table 5.4 (a). The 4 clusters determined by the neighbor-Net splits graph in period 4: 13 January 2004 to 02 January 2007.

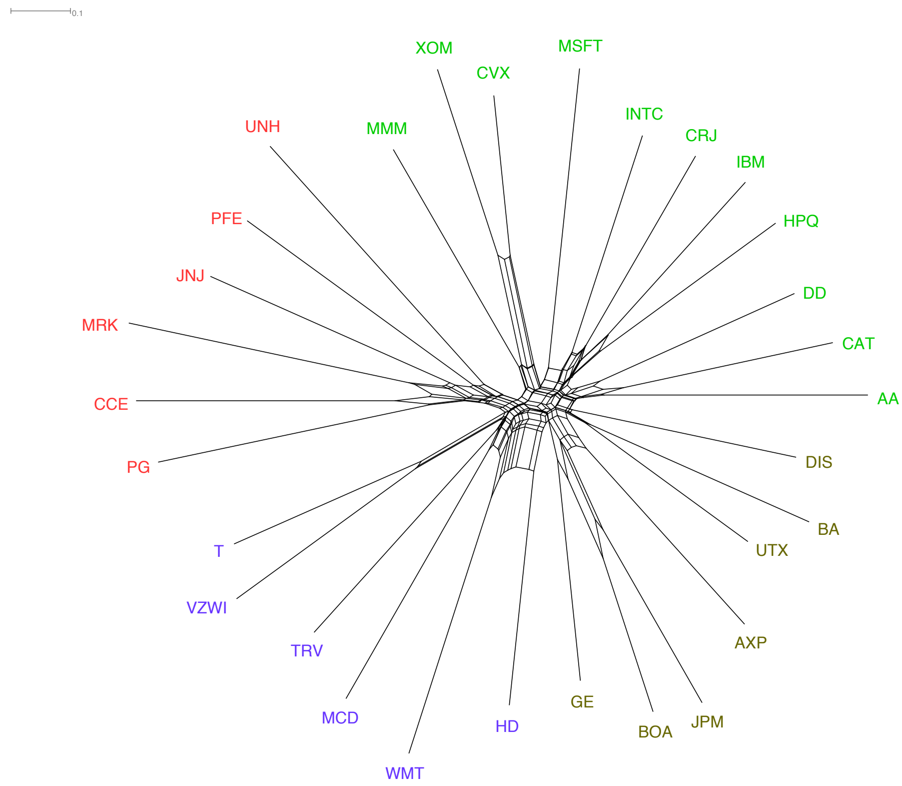


Figure 5.7 (e) Neighbor-Net splits graph produced from stocks' weekly returns in period 5: 09 January 2007 to 05 January 2010.

Cluster 1 (in red)	PG, CCE, MRK, JNJ, PFE, UNH
Cluster 2 (in green)	MMM, XOM, CVX, MSFT, INTC, CRJ, IBM, HPQ, DD, CAT, AA
Cluster 3 (in dark green)	DIS, BA, UTX, AXP, JPM, BOA, GE
Cluster 4 (in purple)	T, VZWI, TRV, MCD, WMT, HD

Table 5.5 (a). The 4 clusters determined by the neighbor-Net splits graph in period 5: 09 January 2007 to 05 January 2010.

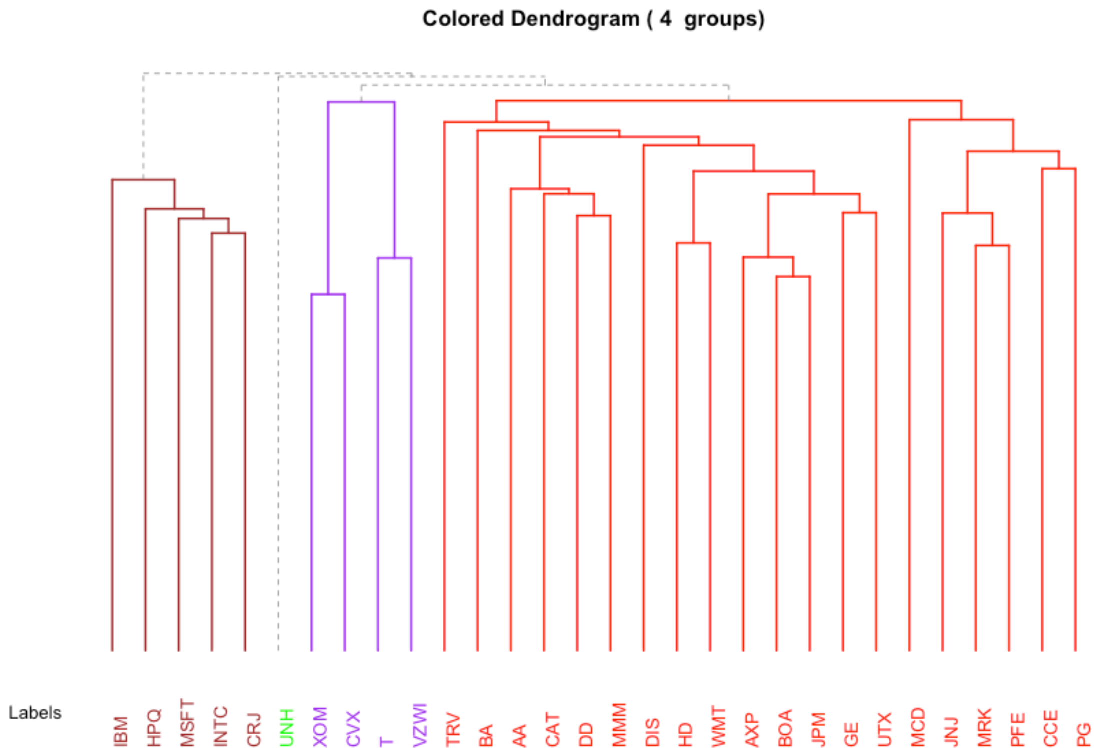


Figure 5.8 (b) The Hierarchical clusters produced from stocks' weekly returns in the period 2: 11 January 1994 to 01 January 2002.

Cluster 1 (in brown)	IBM, HPQ, MSFT, INTC, CRJ
Cluster 2 (in green)	UNH
Cluster 3 (in purple)	XOM, CVX, T, VZWI
Cluster 4 (in red)	TRV, BA, AA, CAT, DD, MMM, DIS, HD, WMT, AXP, BOA, JPM, GE, UTX, MCD, JNJ, MRK, PFE, CCE, PG

Table 5.2 (b). The 4 clusters determined by the HCT in the period 2: 11 January 1994 to 01 January 2002.

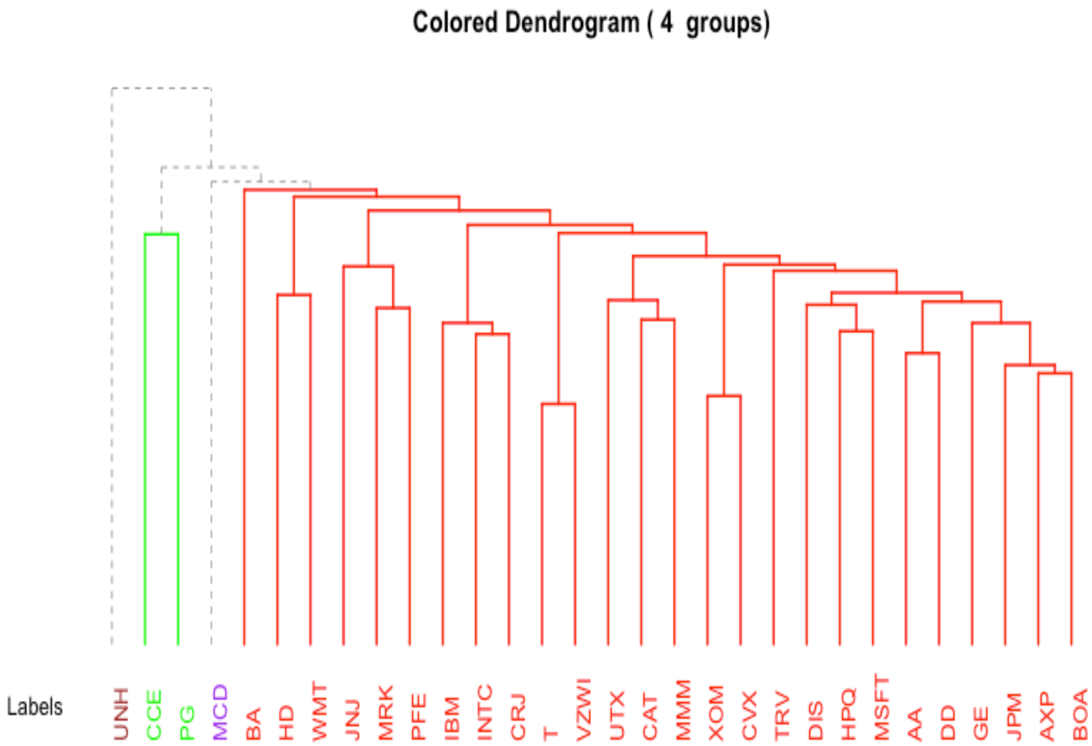


Figure 5.8 (c) The Hierarchical cluster produced from stocks' weekly returns in the period 3: 08 January 2002 to 06 January 2004.

Cluster 1(in brown)	UNH
Cluster 2 (in green)	CCE, PG
Cluster 3 (in purple)	MCD
Cluster 4 (in red)	BA, HD, WMT, JNJ, MRK, PFE, IBM, INTC, CRJ, T, VZWI, UTX, CAT, MMM, XOM, CVX, TRV, DIS, HPQ, MSFT, AA, DD, GE, JPM, AXP, BOA

Table 5.3 (b). The 4 clusters determined by the HCT in the period 3: 08 January 2002 to 06 January 2004.

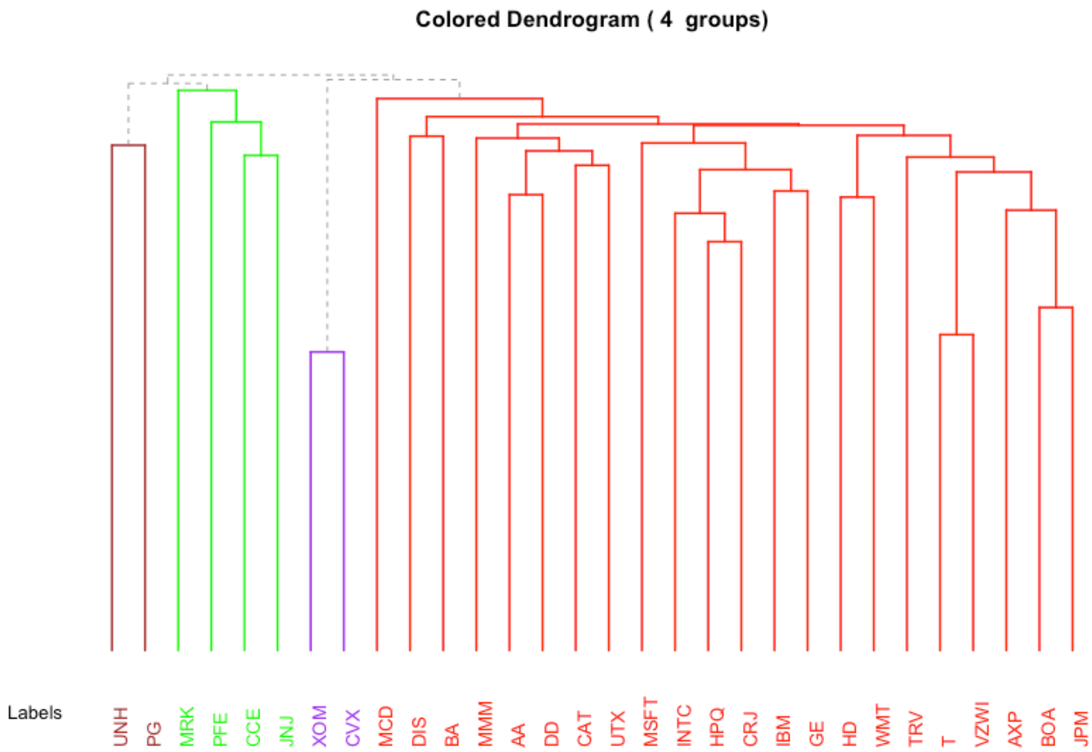


Figure 5.8 (d) The Hierarchical clusters produced from stocks' weekly returns in the period 4: 13 January 2004 to 02 January 2007.

Cluster 1(in brown)	UNH, PG
Cluster 2 (in green)	MRK, PFE, CCE, JNJ
Cluster 3 (in purple)	XOM, CVX
Cluster 4 (in red)	MCD, DIS, BA, MMM, AA, DD, CAT, UTX, MSFT, INTC, HPQ, CRJ, IBM, GE, HD, WMT, TRV, T, VZWI, AXP, BOA, JPM

Table 5.4 (b). The 4 clusters determined by the HCT in the period 4: 13 January 2004 to 02 January 2007.

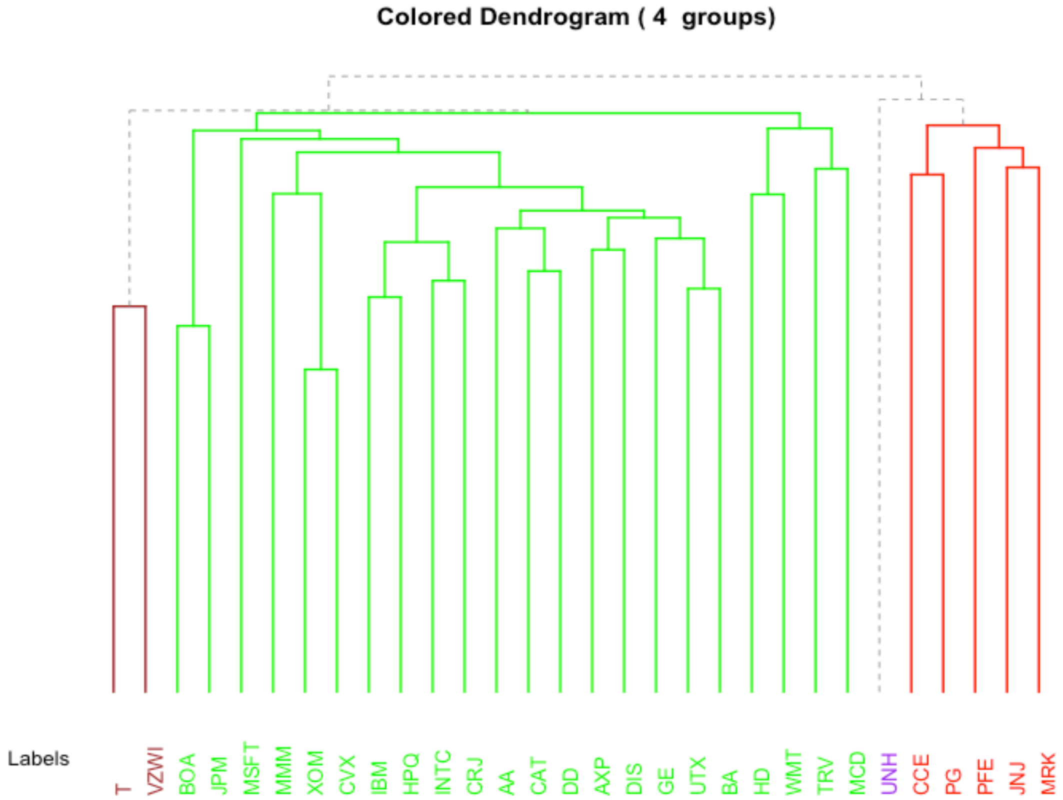


Figure 5.8 (e) The Hierarchical clusters produced from stocks' weekly returns in the period 5: 09 January 2007 to 05 January 2010.

Cluster 1(in brown)	T, VZWI
Cluster 2 (in green)	BOA, JPM, MSFT, MMM, XOM, CVX, IBM, HPQ, INTC, CRJ, AA, CAT, DD, AXP, DIS, GE, UTX, BA, HD, WMT, TRV, MCD
Cluster 3(in purple)	UNH
Cluster 4 (in red)	CCE, PG, PFE, JNJ, MRK

Table 5.5 (b). The 4 clusters determined by the HCT in the period 5: 09 January 2007 to 05 January 2010.

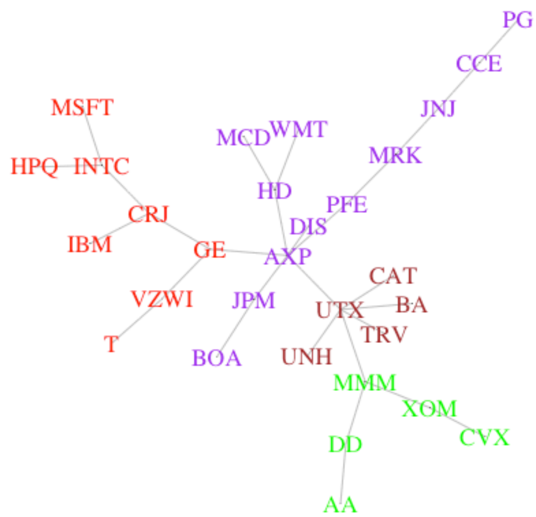


Figure 5.9 (b) MST produced from the stocks' weekly returns in period 2: 11 January 1994 to 01 January 2002.

Cluster 1(in Red)	T, VZWI, GE, CRJ, IBM, INTC, MSFT, HPQ
Cluster 2 (purple)	BOA, JPM, AXP, PFE, MRK, JNJ, CCE, PG, DIS, HD, MCD, WMT
Cluster 3 (in brown)	TRV, CAT, UTX, BA, UNH
Cluster 4 (Green)	MMM, XOM, CVX, DD, AA

Table 5.2 (c) The 4 clusters determined by the MST in period 2: 11 January 1994 to 01 January 2002.

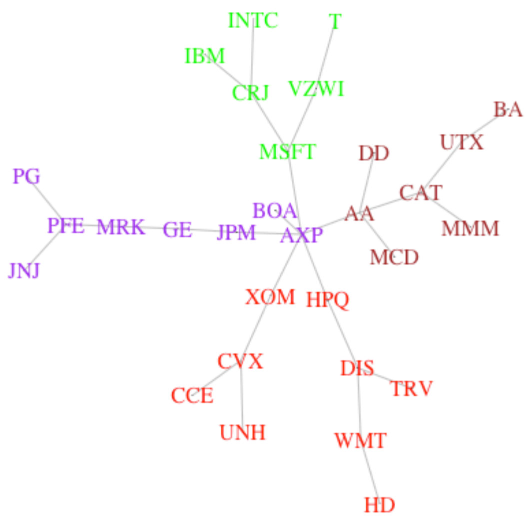


Figure 5.9 (c) MST produced from the stocks' weekly returns in the period 3: 08 January 2002 to 06 January 2004.

Cluster 1(in Red)	HD, WMT, TRV, DIS, HPQ, XOM, CVX, UNH, CCE
Cluster 2 (in brown)	DD, AA, MCD, CAT, MMM, UTX, BA
Cluster 3 (in purple)	AXP, BOA, JPM, GE, MRK, PFE, JNJ, PG
Cluster 4 (in green)	MSFT, VZWI, T, CRJ, INTC, IBM

Table 5.3 (c). The 4 clusters determined by the MST in the period 3: 08 January 2002 to 06 January 2004.

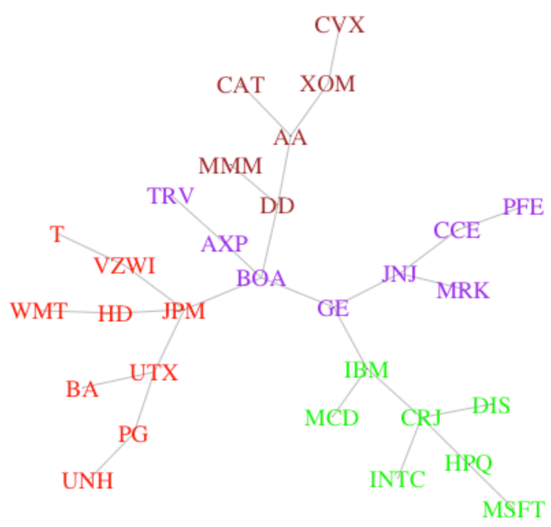


Figure 5.9 (d) MST produced from the stocks' weekly returns in period 4: 13 January 2004 to 02 January 2007.

Cluster 1(in red)	UNH, PG, UTX, BA, JPM, VZWI, T, HD, WMT
Cluster 2 (in brown)	DD, MMM, AA, CAT,XOM,CVX
Cluster 3 (in purple)	TRV, AXP, BOA, GE, JNJ, MRK, CCE, PFE
Cluster 4 (in green)	IBM, MCD, CRJ, DIS, INTC, HPQ, MSFT

Table 5.4 (c). The 4 clusters determined by the MST in period 4: 13 January 2004 to 02 January 2007.

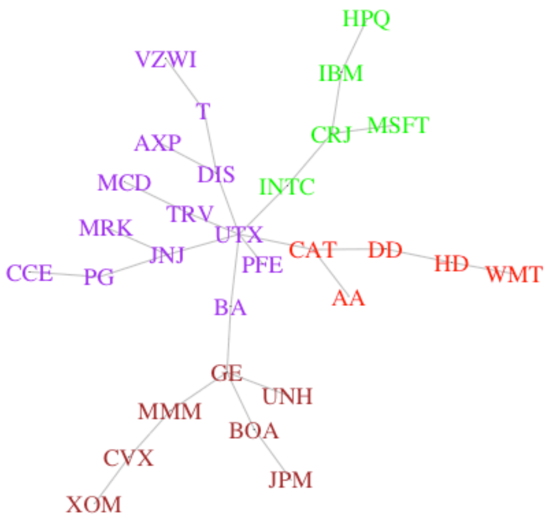


Figure 5.9 (e) MST produced from the stocks' weekly returns in period 5: 09 January 2007 to 05 January 2010.

Cluster 1(in brown)	JPM, BOA, GE, UNH, MMM, CVX, XOM
Cluster 2 (in purple)	VZWI, T, AXP, DIS, UTX, PFE, BA, MCD, TRV, MRK, JNJ, PG, CCE
Cluster 3 (in red)	CAT, AA, DD, HD, WMT
Cluster 4 (in green)	INTC, CRJ, MSFT, IBM, HPQ

Table 5.5 (c). The 4 clusters determined by the MST in period 5: 09 January 2007 to 05 January 2010.

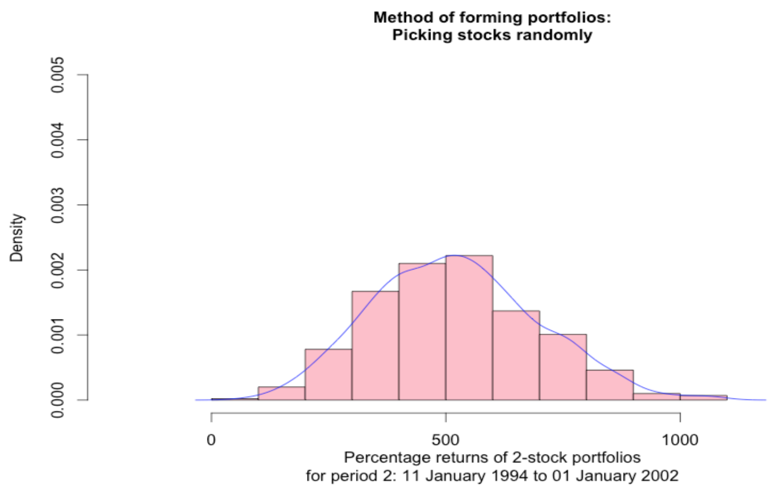


Figure 5.10 (a). The histogram and density curve of the simulated portfolios' percentage returns for period 2: 11 January 1994 to 01 January 2002. Each portfolio contains two stocks which were picked randomly.

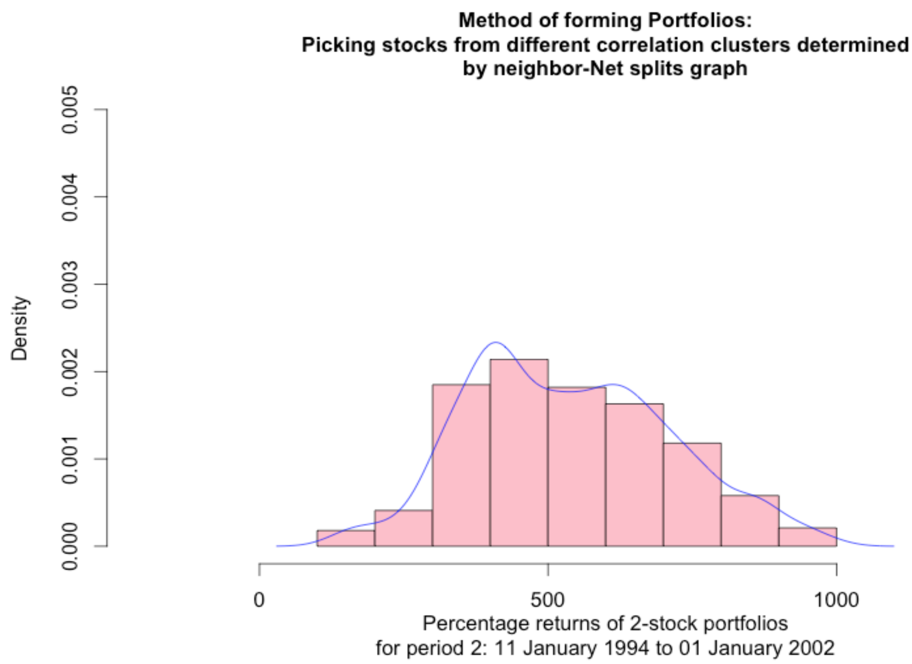


Figure 5.10 (b). The histogram and density curve of the simulated portfolios' percentage returns for period 2: 11 January 1994 to 01 January 2002. Each portfolio contains two stocks which were picked from different correlation clusters suggested by the neighbor-Net split graph produced from the stocks' weekly returns in period 1: 20 February 1990 to 04 January 1994.

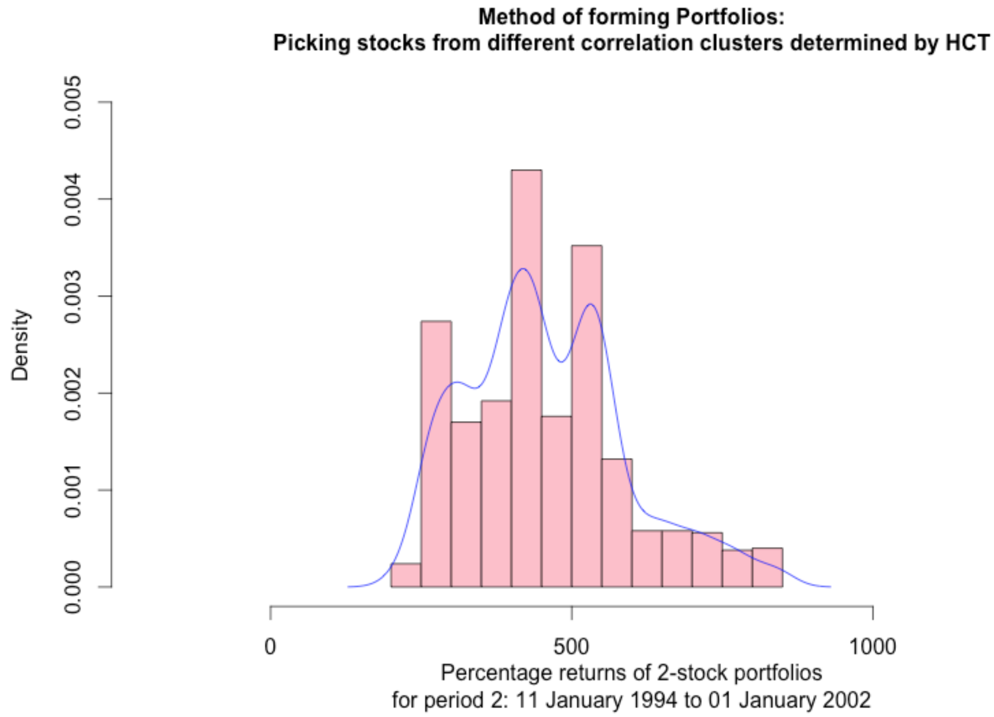


Figure 5.10 (c). The histogram and density curve of the simulated portfolios' percentage returns for period 2: 11 January 1994 to 01 January 2002. Each portfolio contains two stocks which were picked from different correlation clusters suggested by the HCT produced from stocks' weekly returns period 1: 20 February 1990 to 04 January 1994.

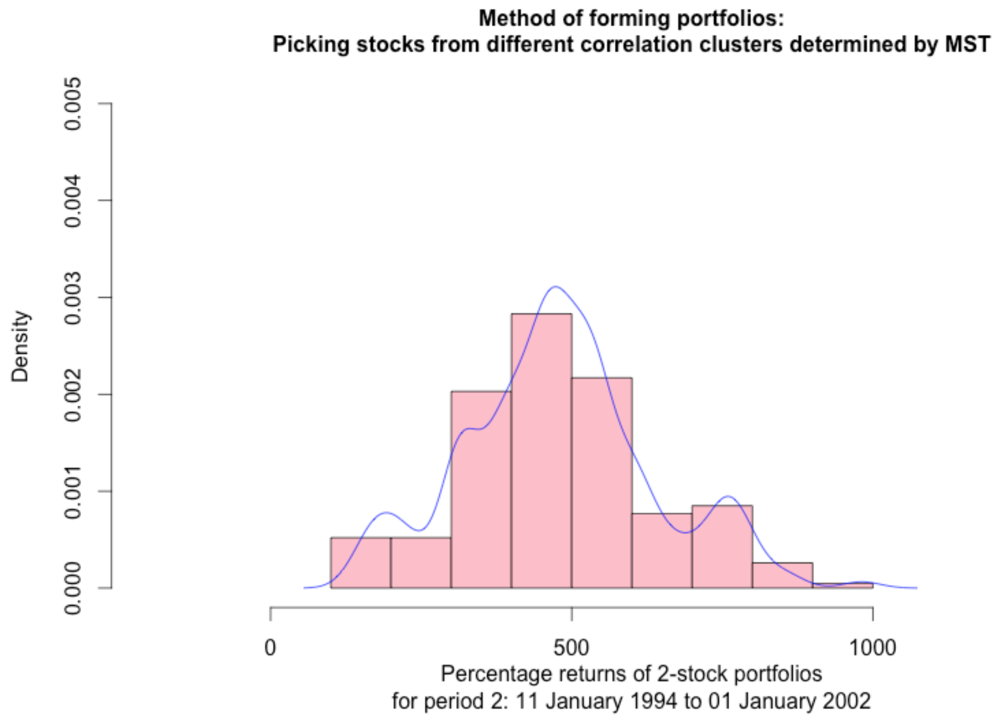


Figure 5.10 (d). The histogram and density curve of the simulated portfolios' percentage returns for period 2: 11 January 1994 to 01 January 2002. Each portfolio contains two stocks which were picked from different correlation clusters suggested by the MST produced from the stocks' weekly returns in period 1: 20 February 1990 to 04 January 1994.

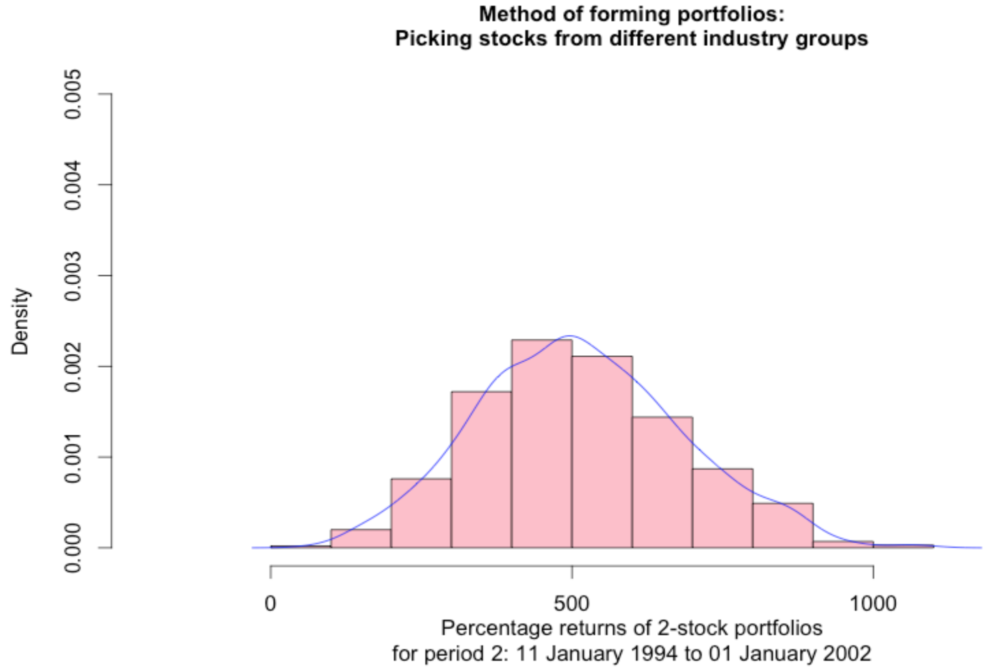


Figure 5.10 (e). The histogram and density curve of the simulated portfolios' percentage returns for period 2: 11 January 1994 to 01 January 2002. Each portfolio contains two stocks which were picked from different industry groups.

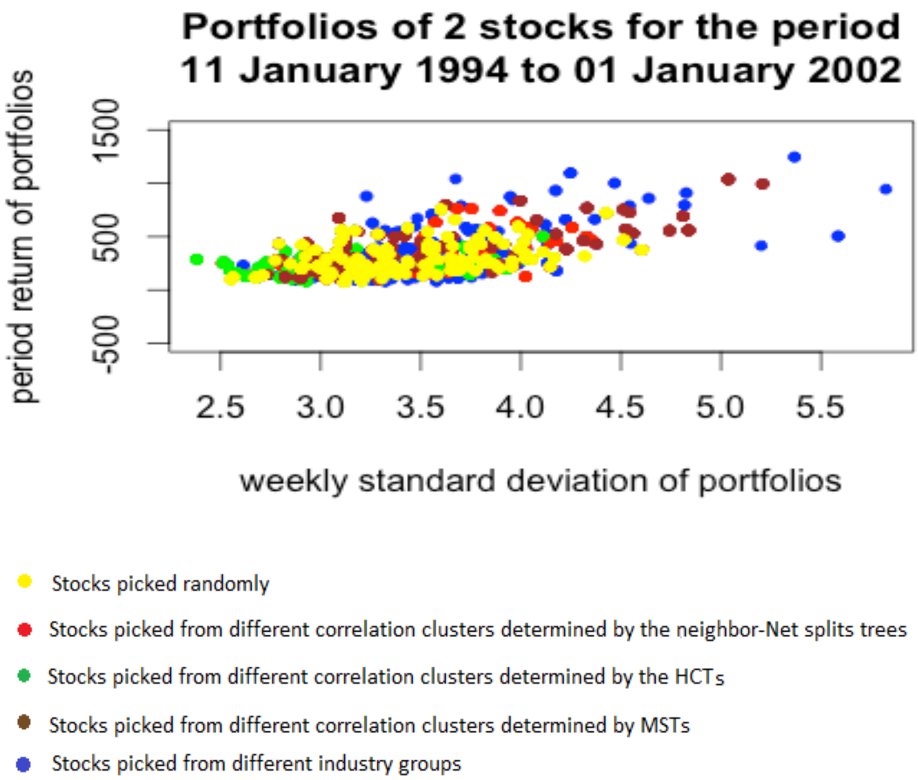


Figure 5.10 (f). Weekly standard deviation and return of each set of simulated 2-stock portfolios for period 2 using the five methods. The period returns and weekly standard deviations are expressed in percentage.

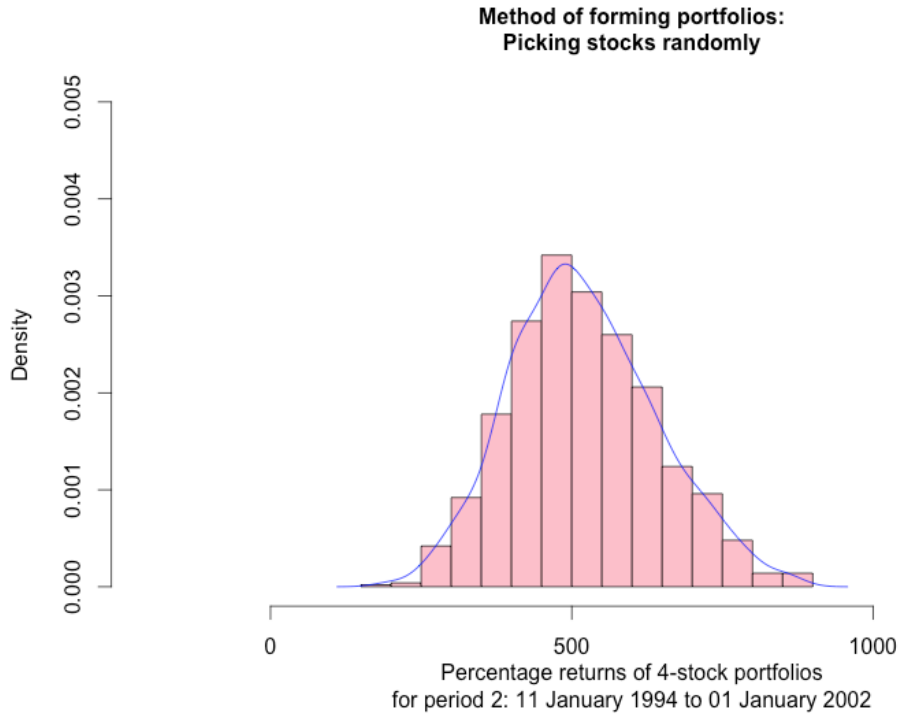


Figure 5.11 (a). The histogram and density curve of the simulated portfolios' percentage returns for period 2: 11 January 1994 to 01 January 2002. Each portfolio contains four stocks which were picked randomly.

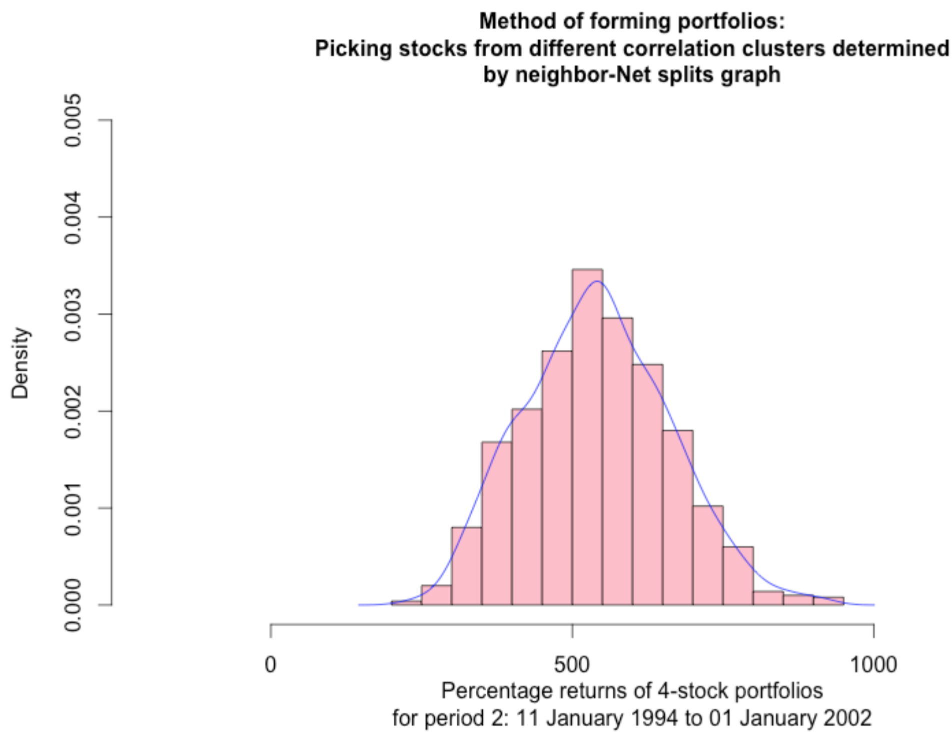


Figure 5.11 (b). The histogram and density curve of the simulated portfolios' percentage returns for period 2: 11 January 1994 to 01 January 2002. Each portfolio contains four stocks which were picked from different correlation clusters suggested by the neighbor-Net split graph produced from the stocks weekly returns in period 1: 20 February 1990 to 04 January 1994.

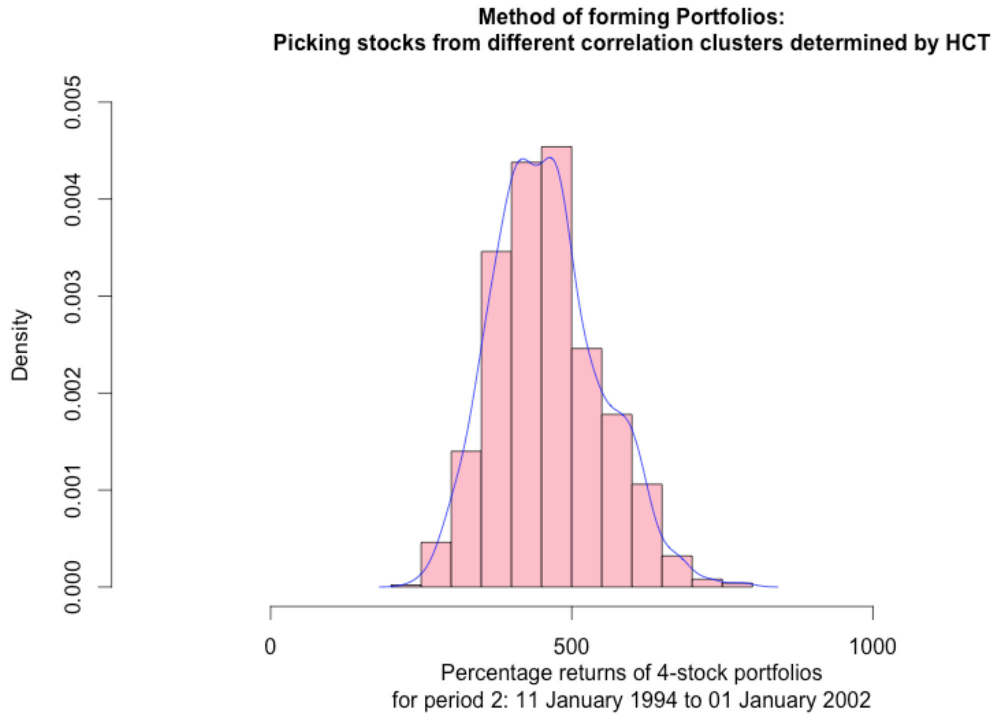


Figure 5.11 (c). The histogram and density curve of the simulated portfolios' percentage returns for period 2:11 January 1994 to 01 January 2002. Each portfolio contains four stocks which were picked from different correlation clusters suggested by the HCT produced from stocks weekly returns in period 1: 20 February 1990 to 04 January 1994.

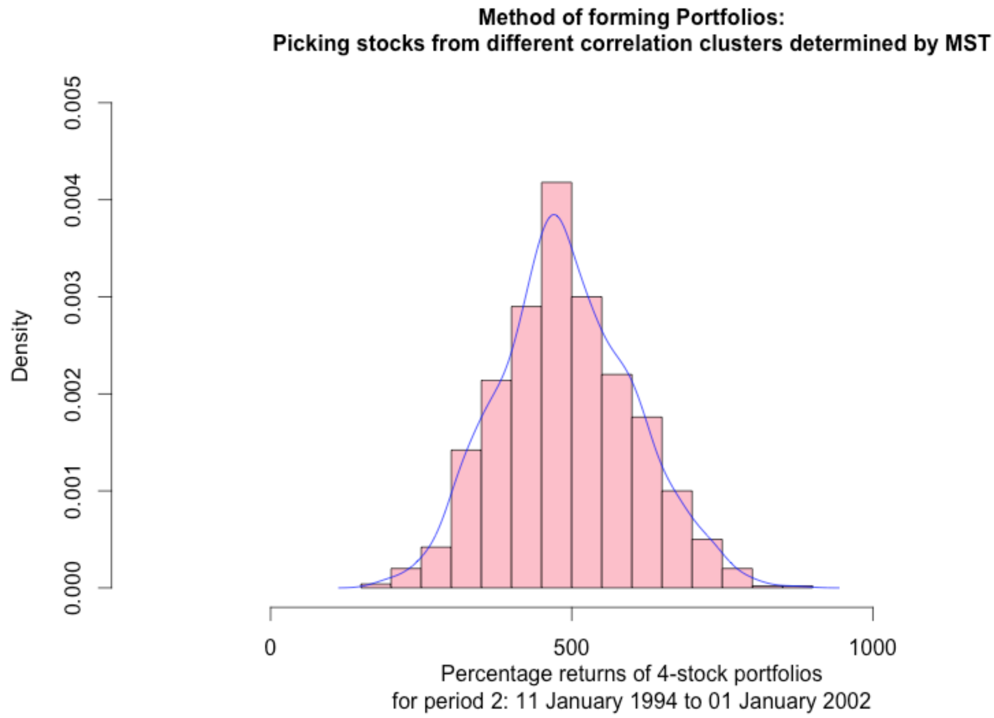


Figure 5.11 (d). The histogram and density curve of the simulated portfolios' percentage returns for period 2: 11 January 1994 to 01 January 2002. Each portfolio contains four stocks which were picked from different correlation clusters suggested by the MST produced from the stocks weekly returns in period 1: 20 February 1990 to 04 January 1994.

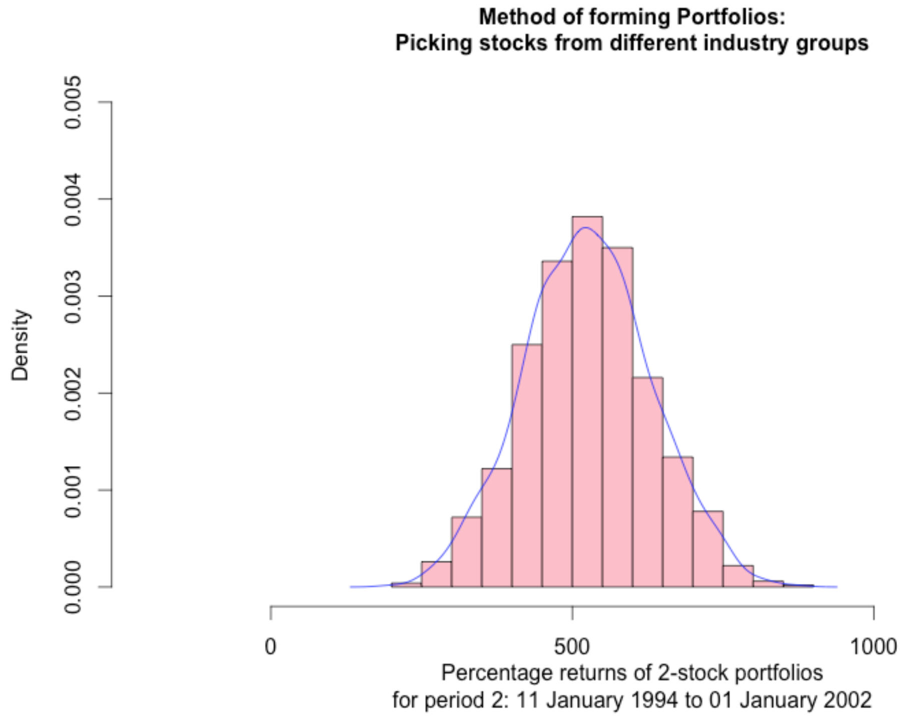


Figure 5.11 (e). The histogram and density curve of the simulated portfolios' percentage returns for period 2: 11 January 1994 to 01 January 2002. Each portfolio contains four stocks which were picked from different industry groups.

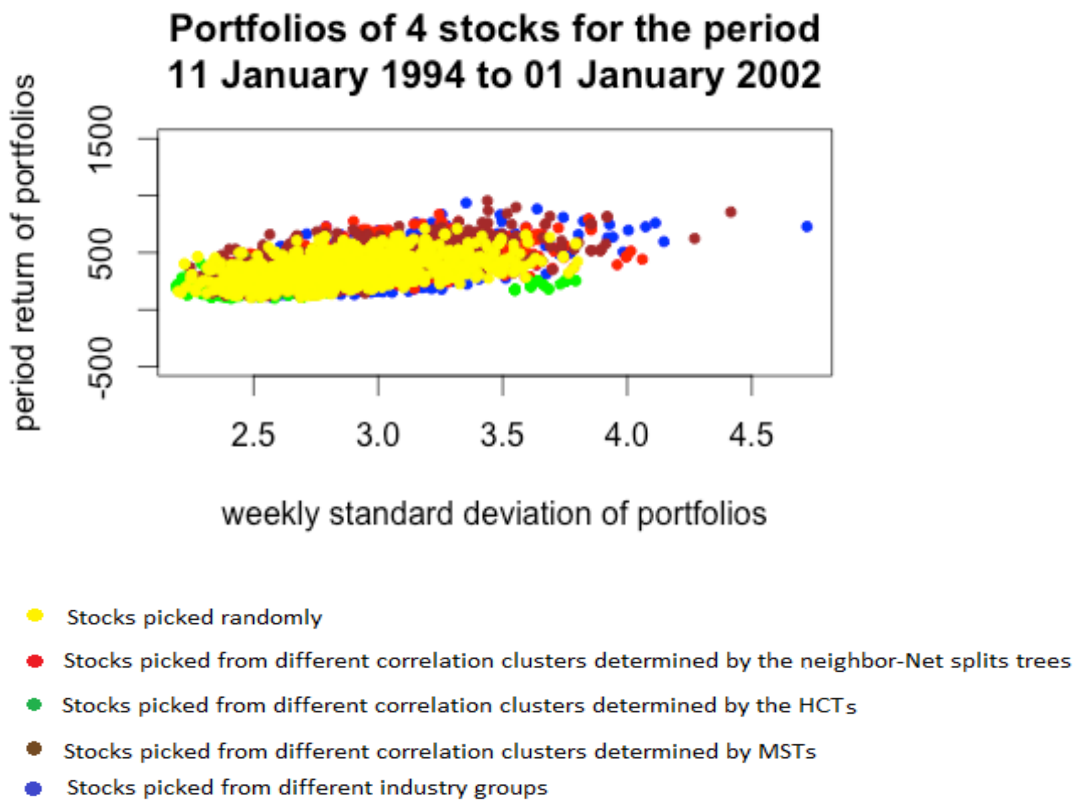


Figure 5.11 (f). Weekly standard deviation and return of each set of simulated 4-stock portfolios for period 2 using the five methods. The period returns and weekly standard deviations are expressed in percentage.

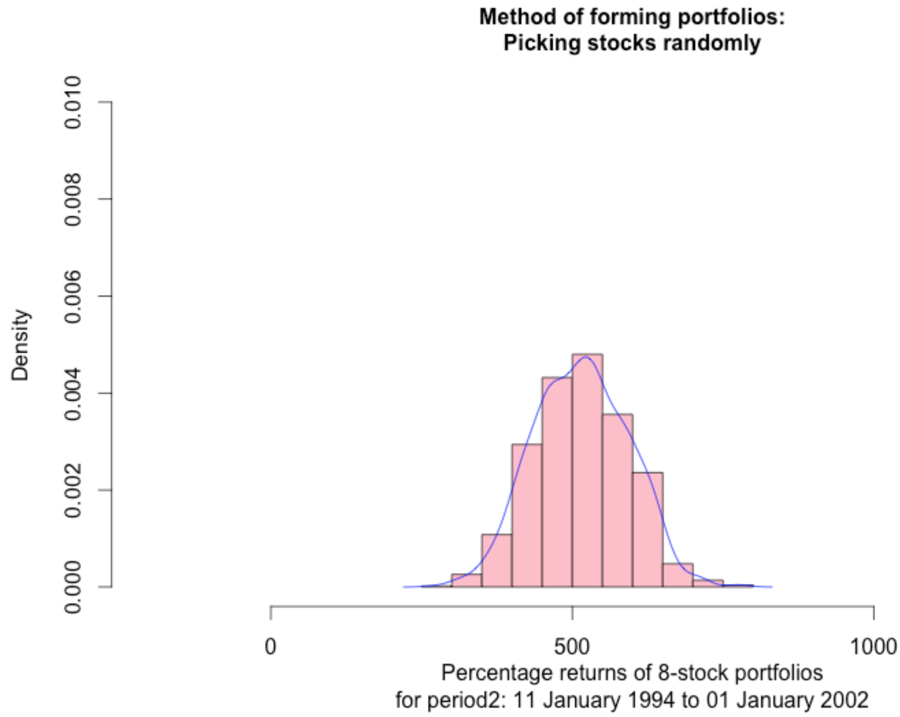


Figure 5.12 (a). The histogram and density curve of the simulated portfolios' percentage returns for period 2: 11 January 1994 to 01 January 2002. Each portfolio contains eight stocks which were picked randomly.

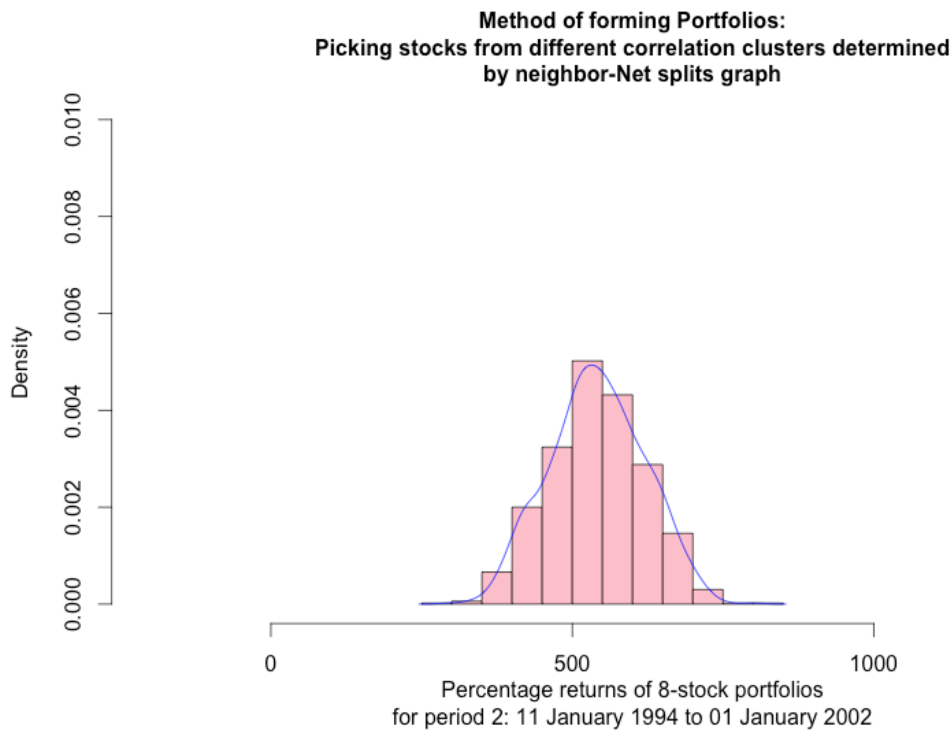


Figure 5.12 (b). The histogram and density curve of the simulated portfolios' percentage returns for period 2: 11 January 1994 to 01 January 2002. Each portfolio contains eight stocks which were picked from different correlation clusters suggested by the neighbor-Net split graph produced from the stocks weekly returns in period 1: 20 February 1990 to 04 January 1994.

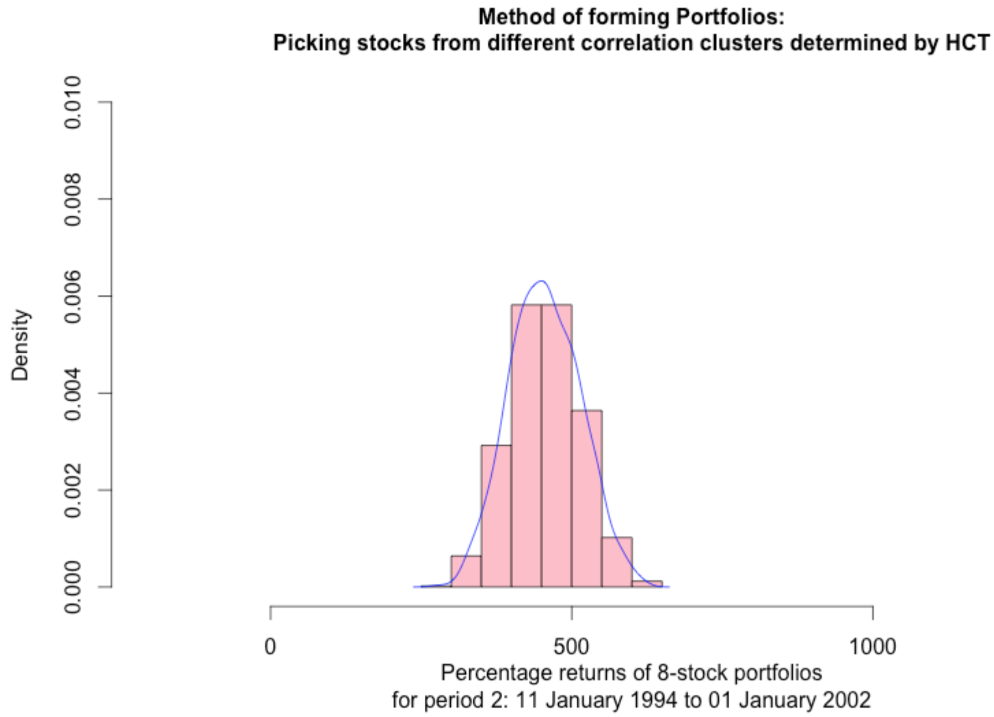


Figure 5.12(c). The histogram and density curve of the simulated portfolios' percentage returns for period 2: 11 January 1994 to 01 January 2002. Each portfolio contains eight stocks which were picked from different correlation clusters suggested by the HCT produced from stocks weekly returns in period 1: 20 February 1990 to 04 January 1994.

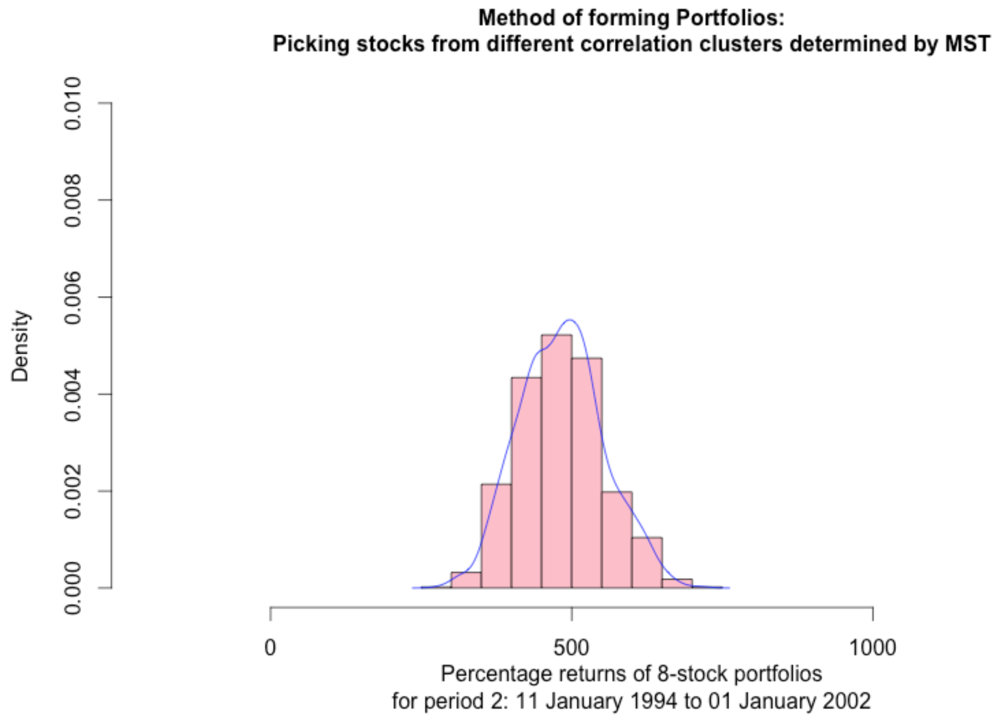


Figure 5.12 (d). The histogram and density curve of the simulated portfolios' percentage returns for period 2: 11 January 1994 to 01 January 2002. Each portfolio contains eight stocks which were picked from different correlation clusters suggested by the MST produced from the stocks weekly returns in period 1: 20 February 1990 to 04 January 1994.

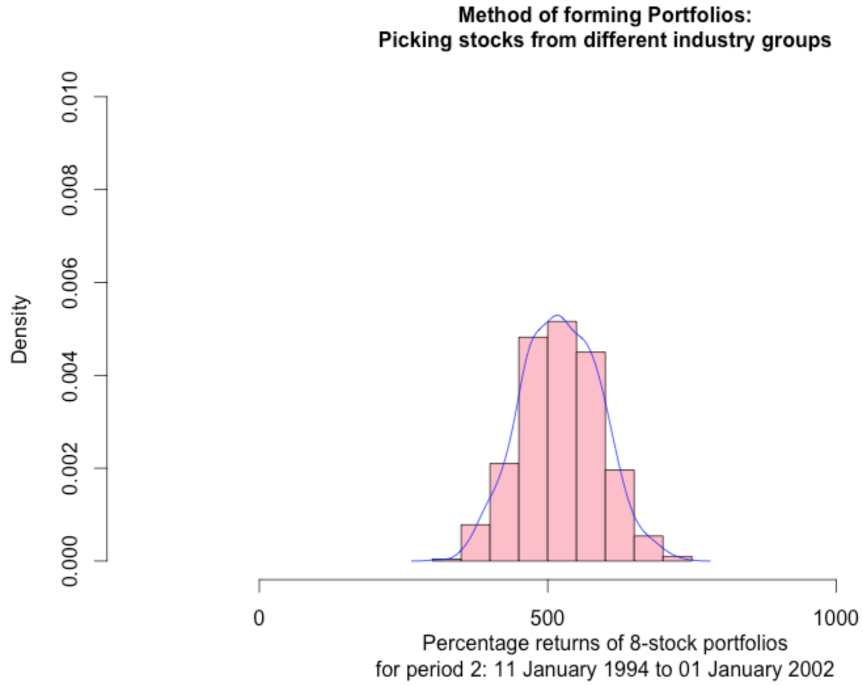


Figure 5.12 (e). The histogram and density curve of the simulated portfolios' percentage returns for period 2: 11 January 1994 to 01 January 2002. Each portfolio contains eight stocks which were picked from different industry groups.

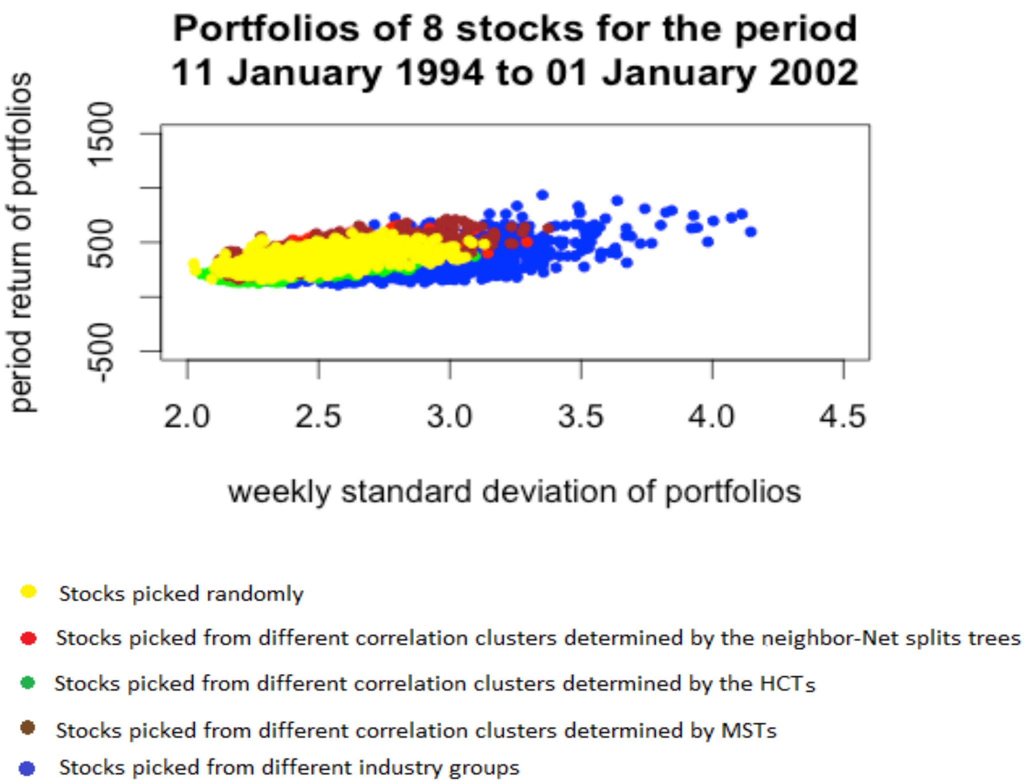


Figure 5.12 (f) Weekly standard deviation and return of each set of simulated 8-stock portfolios for period 2 using the five methods. The period returns and weekly standard deviations are expressed in percentage.

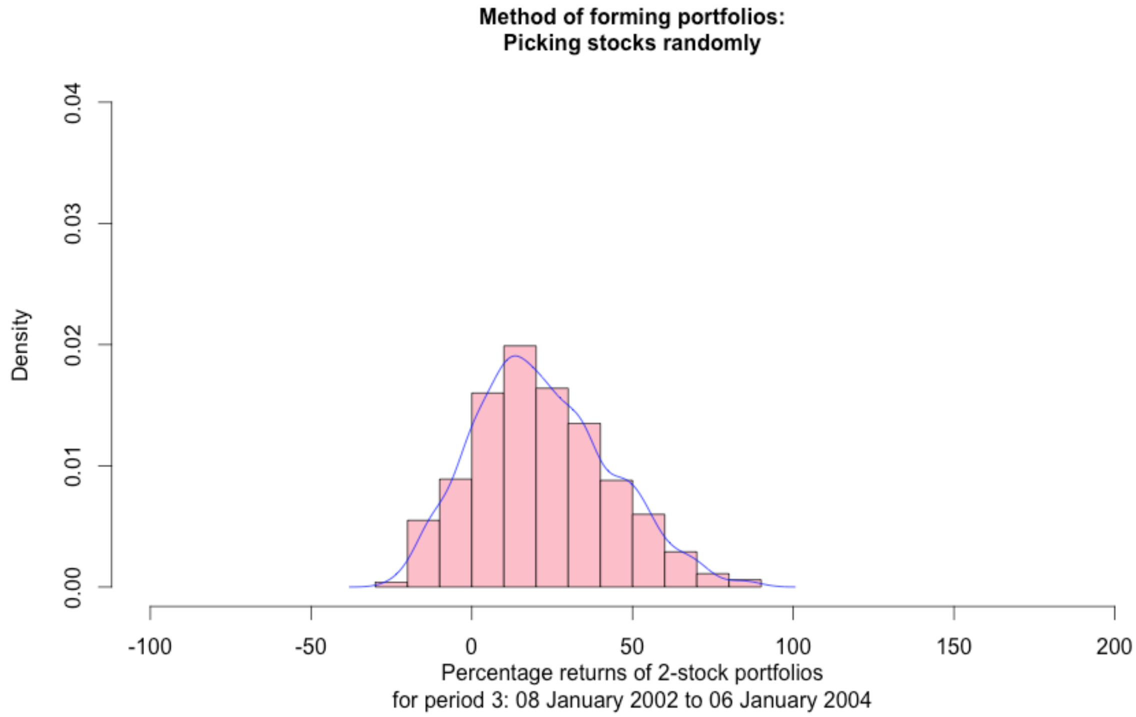


Figure 5.13 (a). The histogram and density curve of the simulated portfolios' percentage returns for period 3: 08 January 2002 to 06 January 2004. Each portfolio contains two stocks which were picked randomly.

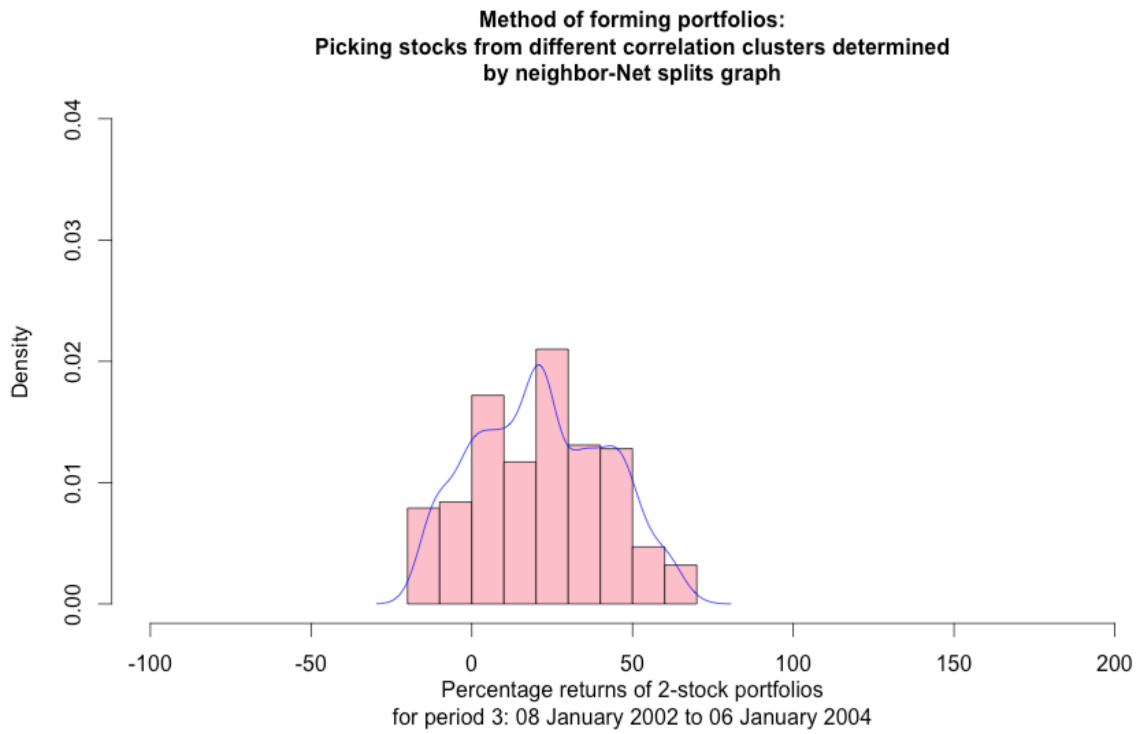


Figure 5.13 (b). The histogram and density curve of the simulated portfolios' percentage returns for period 3: 08 January 2002 to 06 January 2004. Each portfolio contains two stocks which were picked from different correlation clusters suggested by the neighbor-Net split graph produced from the stocks weekly returns in period 2: 11 January 1994 to 01 January 2002.

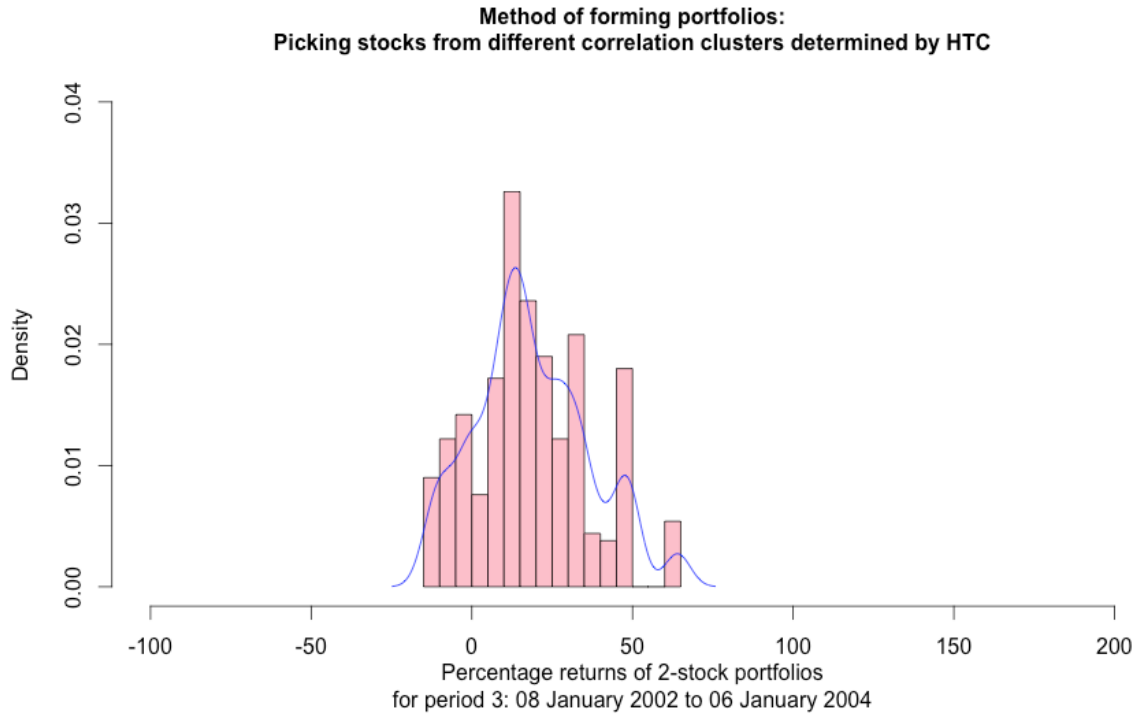


Figure 5.13 (c). The histogram and density curve of the simulated portfolios' percentage returns for period 3: 08 January 2002 to 06 January 2004. Each portfolio contains two stocks which were picked from different correlation clusters suggested by the HCT produced from the stocks weekly returns in period 2: 11 January 1994 to 01 January 2002.

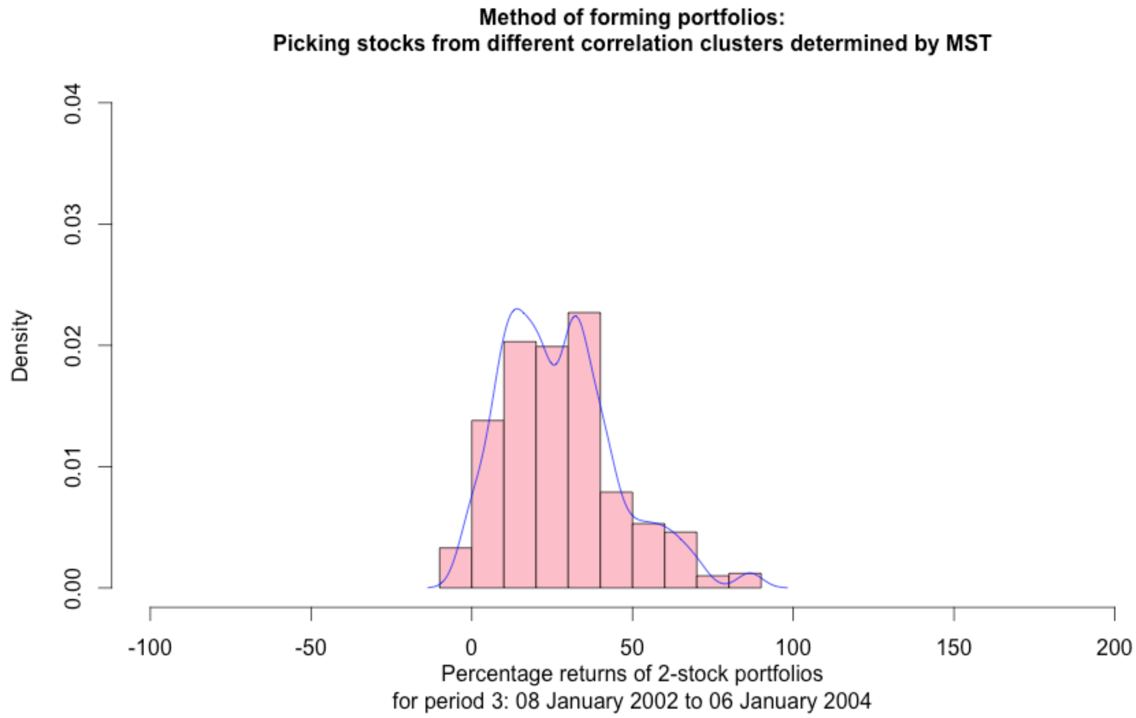


Figure 5.13 (d). The histogram and density curve of the simulated portfolios' percentage returns for period 3: 08 January 2002 to 06 January 2004. Each portfolio contains two stocks which were picked from different correlation clusters suggested by the MST produced from the stocks weekly returns in period 2: 11 January 1994 to 01 January 2002.

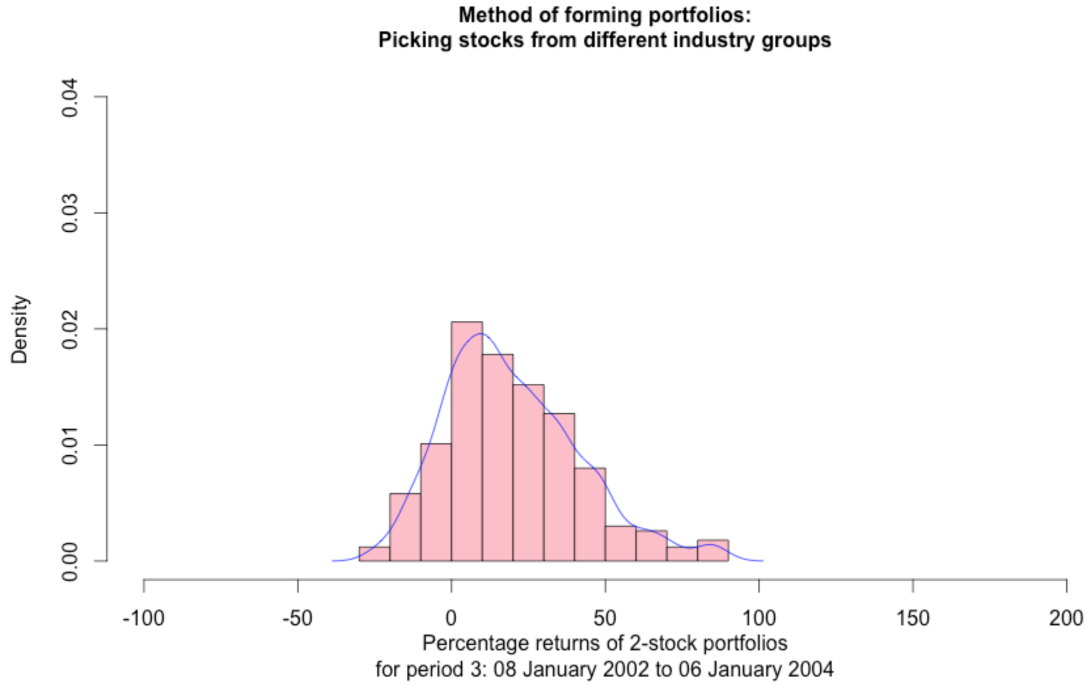


Figure 5.13 (e). The histogram and density curve of the simulated portfolios' percentage returns for period 3: 08 January 2002 to 06 January 2004. Each portfolio contains two stocks which were picked from different industry groups.

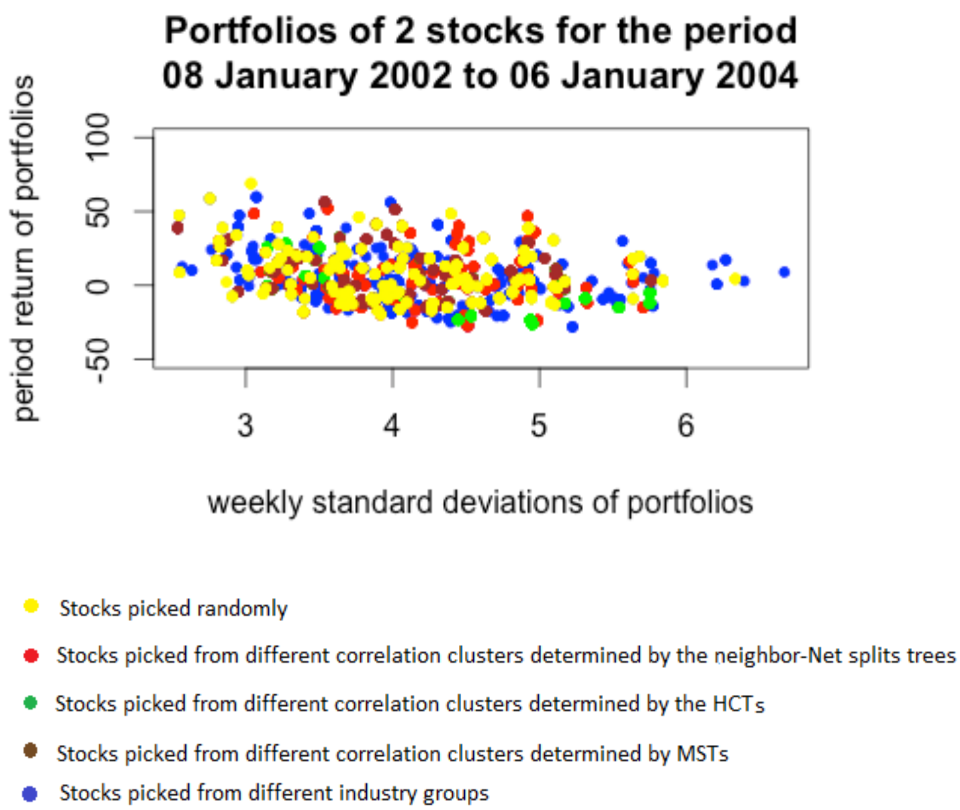


Figure 5.13 (f) Weekly standard deviation and return of each set of simulated 2-stock portfolios for period 3 using the five methods. The period returns and weekly standard deviations are expressed in percentage.

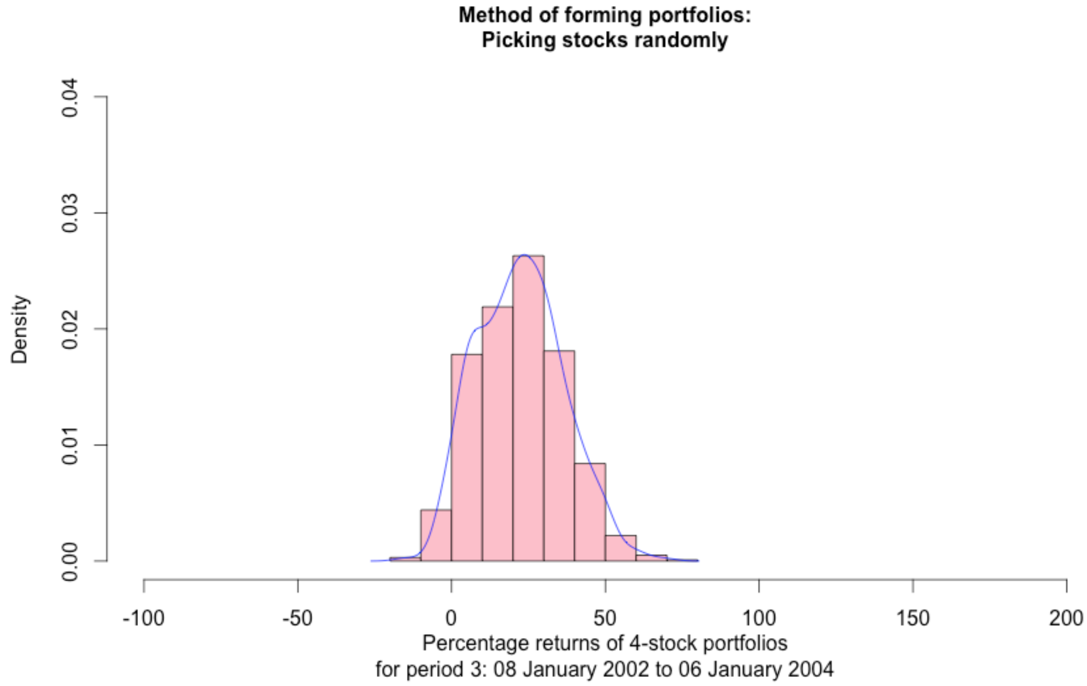


Figure 5.14(a). The histogram and density curve of the simulated portfolios' percentage returns for period 3: 08 January 2002 to 06 January 2004. Each portfolio contains four stocks which were picked randomly.

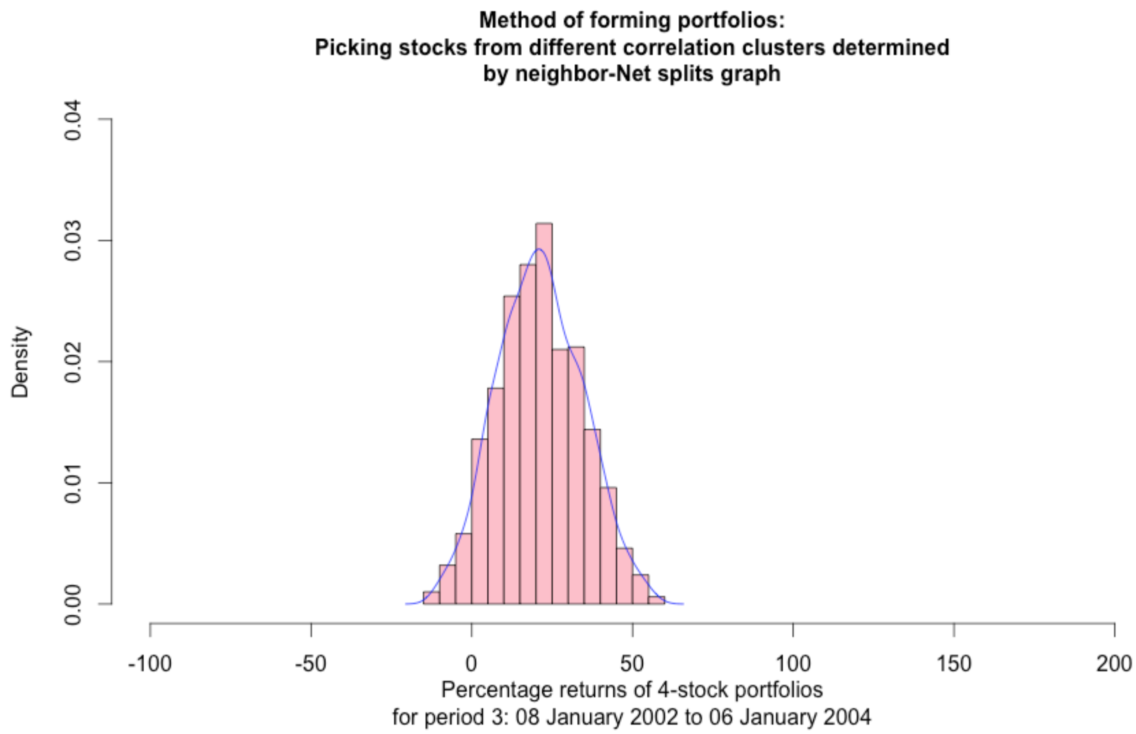


Figure 5.14 (b). The histogram and density curve of the simulated portfolios' percentage returns for period 3: 08 January 2002 to 06 January 2004. Each portfolio contains four stocks which were picked from different correlation clusters suggested by the neighbor-Net split graph produced from the stocks weekly returns in period 2: 11 January 1994 to 01 January 2002.

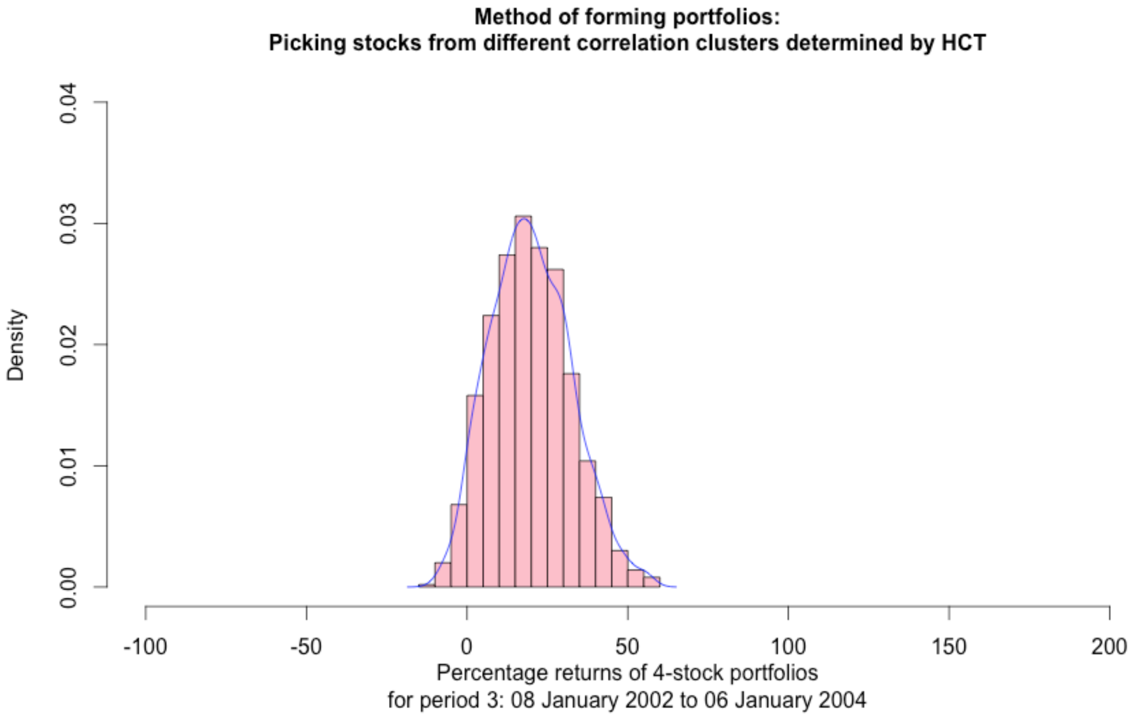


Figure 5.14 (c). The histogram and density curve of the simulated portfolios' percentage returns for period 3: 08 January 2002 to 06 January 2004. Each portfolio contains four stocks which were picked from different correlation clusters suggested by the HCT produced from the stocks weekly returns in period 2: 11 January 1994 to 01 January 2002.

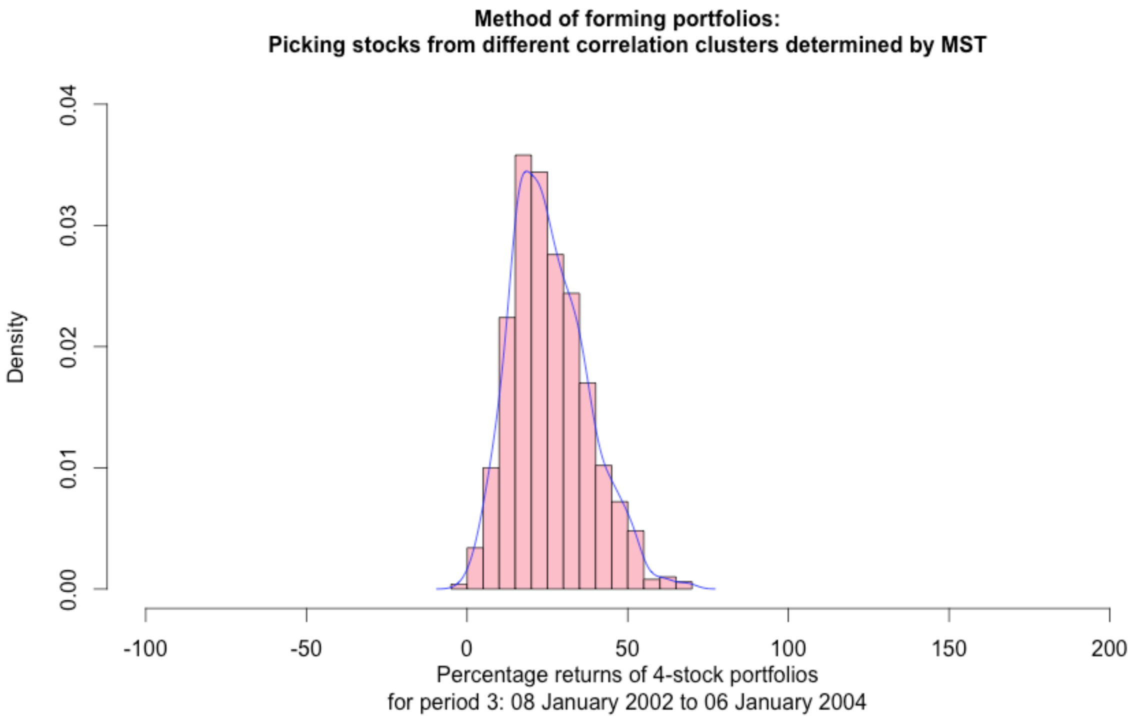


Figure 5.14 (d). The histogram and density curve of the simulated portfolios' percentage returns for period 3: 08 January 2002 to 06 January 2004. Each portfolio contains four stocks which were picked from different correlation clusters suggested by the MST produced from the stocks weekly returns in period 2: 11 January 1994 to 01 January 2002.

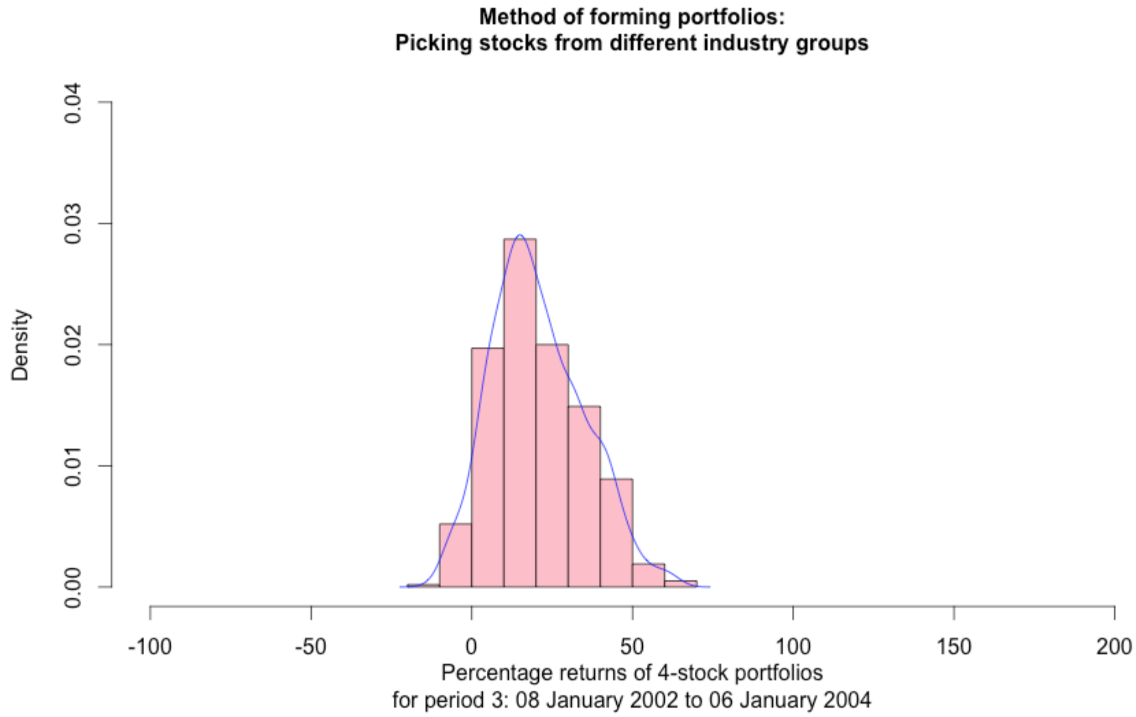


Figure 5.14 (e). The histogram and density curve of the simulated portfolios' percentage returns for period 3: 08 January 2002 to 06 January 2004. Each portfolio contains four stocks which were picked from different industry groups.

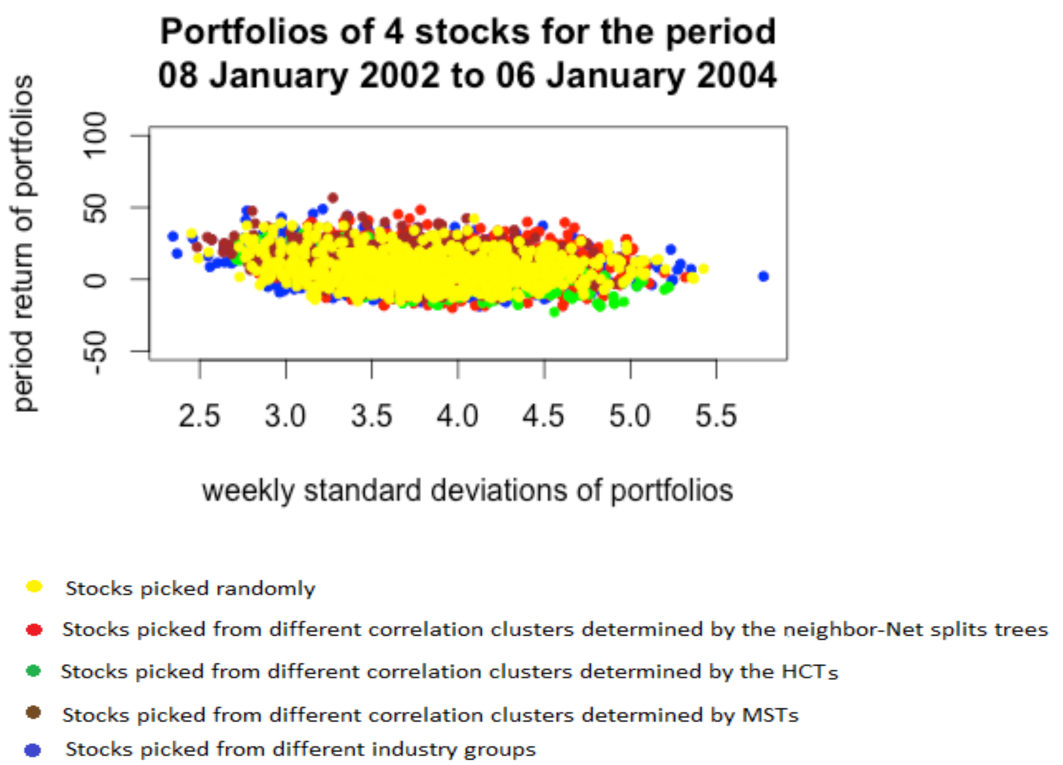


Figure 5.14 (f) Weekly standard deviation and return of each set of simulated 4-stock portfolios for period 3 using the five methods. The period returns and weekly standard deviations are expressed in percentage.

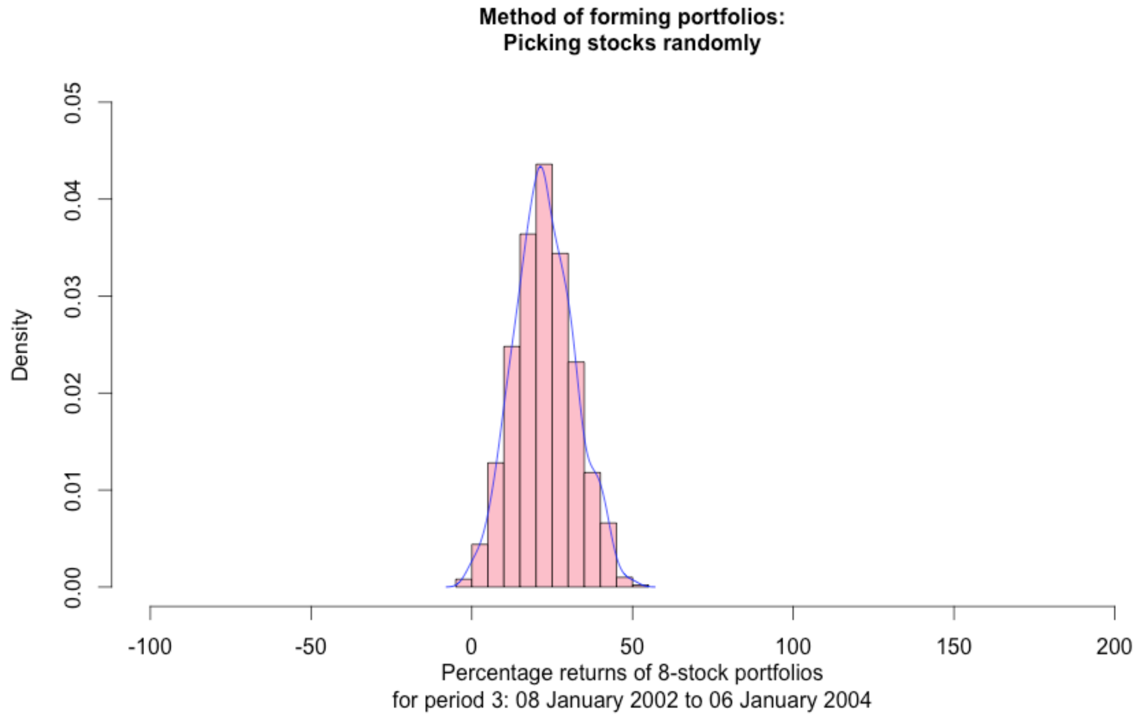


Figure 5.15 (a). The histogram and density curve of the simulated portfolios' percentage returns for period 3: 08 January 2002 to 06 January 2004. Each portfolio contains eight stocks which were picked randomly.

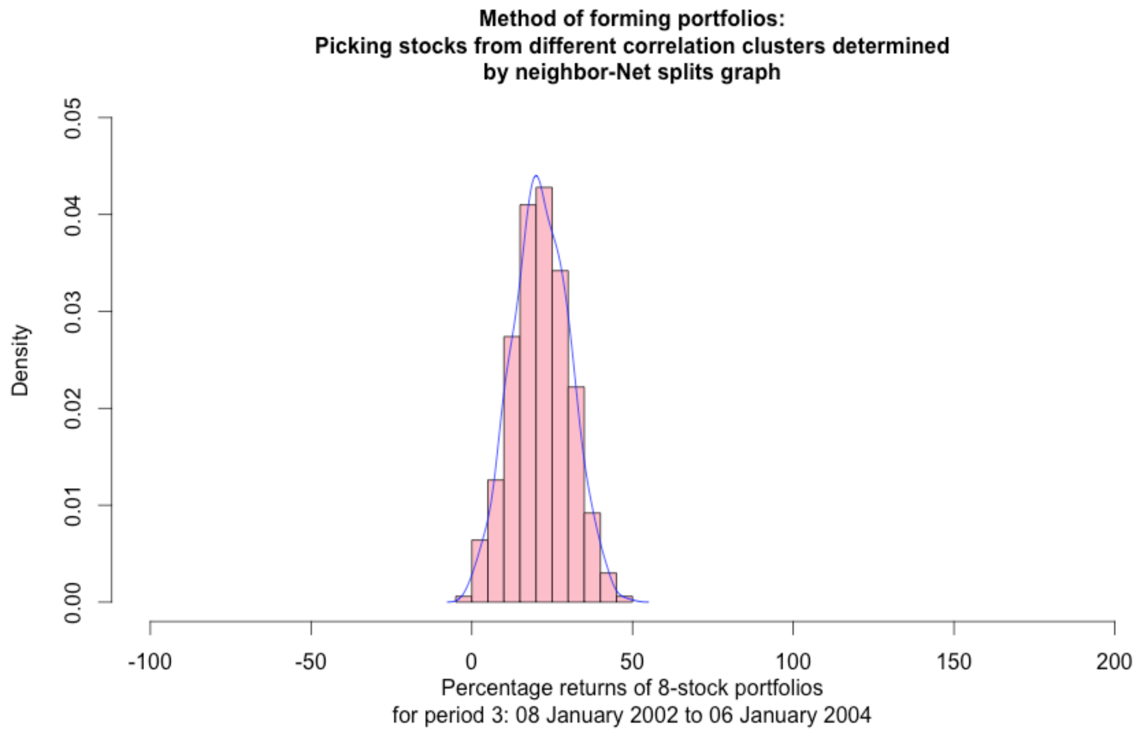


Figure 5.15 (b). The histogram and density curve of the simulated portfolios' percentage returns for period 3: 08 January 2002 to 06 January 2004. Each portfolio contains eight stocks which were picked from different correlation clusters suggested by the neighbor-Net split graph produced from the stocks weekly returns in period 2: 11 January 1994 to 01 January 2002.

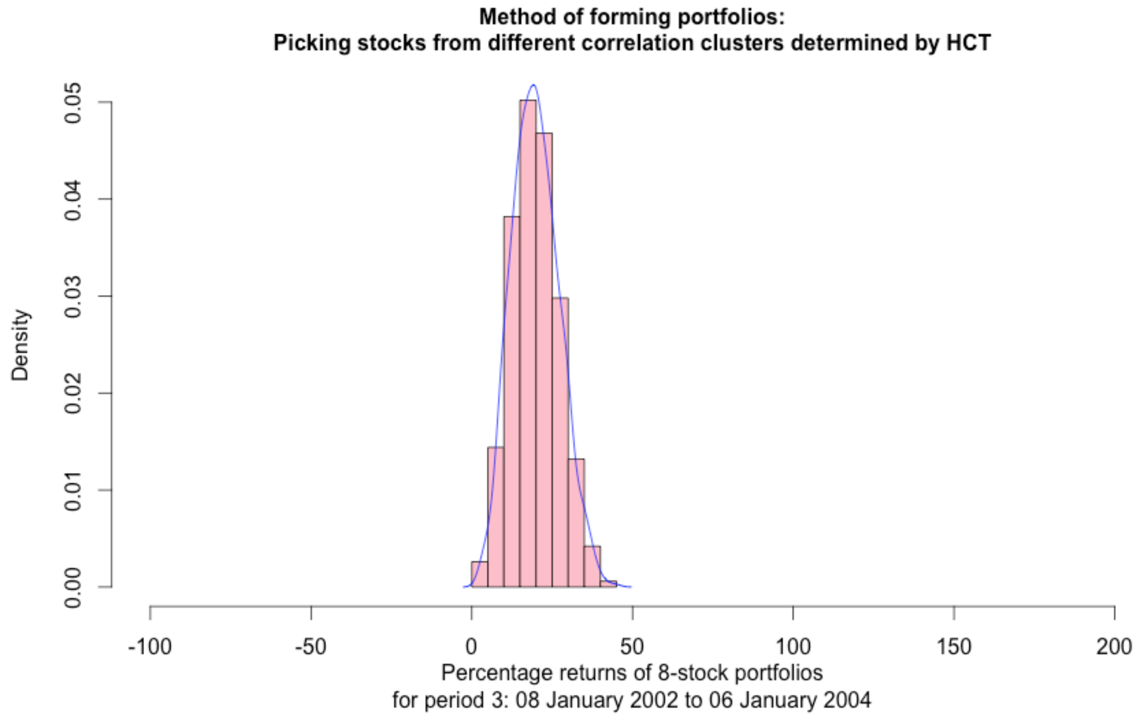


Figure 5.15 (c). The histogram and density curve of the simulated portfolios' percentage returns for period 3: 08 January 2002 to 06 January 2004. Each portfolio contains eight stocks which were picked from different correlation clusters suggested by the HCT produced from the stocks weekly returns in period 2: 11 January 1994 to 01 January 2002.

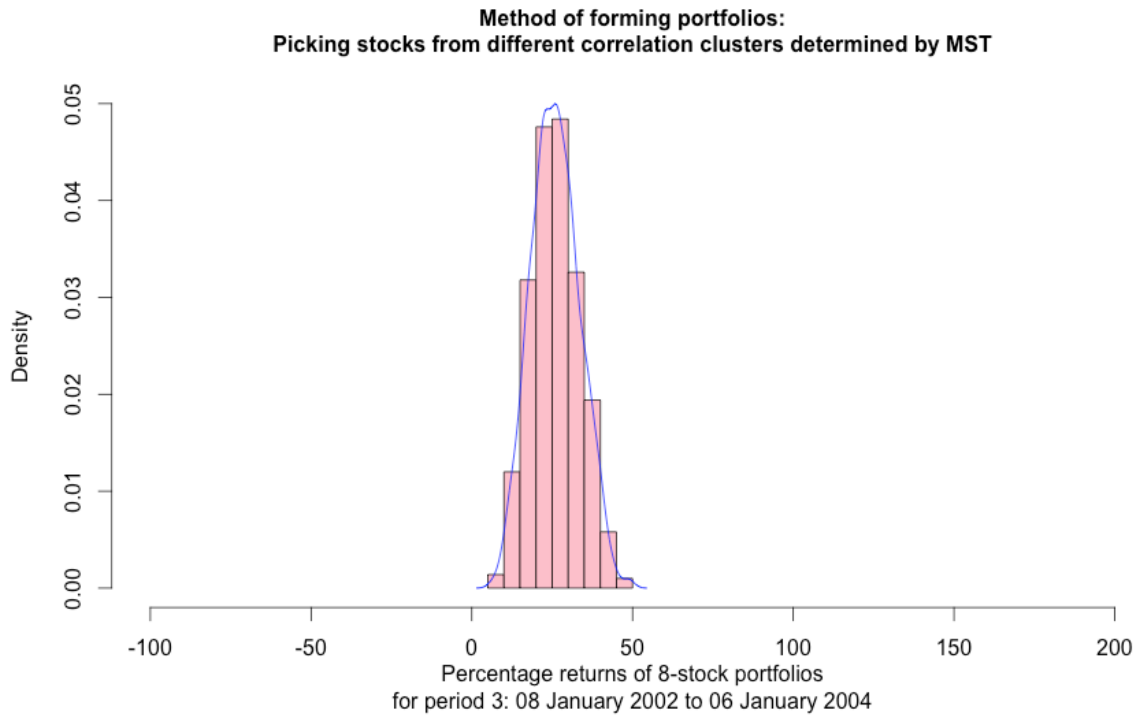


Figure5.15 (d). The histogram and density curve of the simulated portfolios' percentage returns for period 3:08 January 2002 to 06 January 2004. Each portfolio contains eight stocks which were picked from different correlation clusters suggested by the MST produced from the stocks weekly returns in period 2: 11 January 1994 to 01 January 2002.

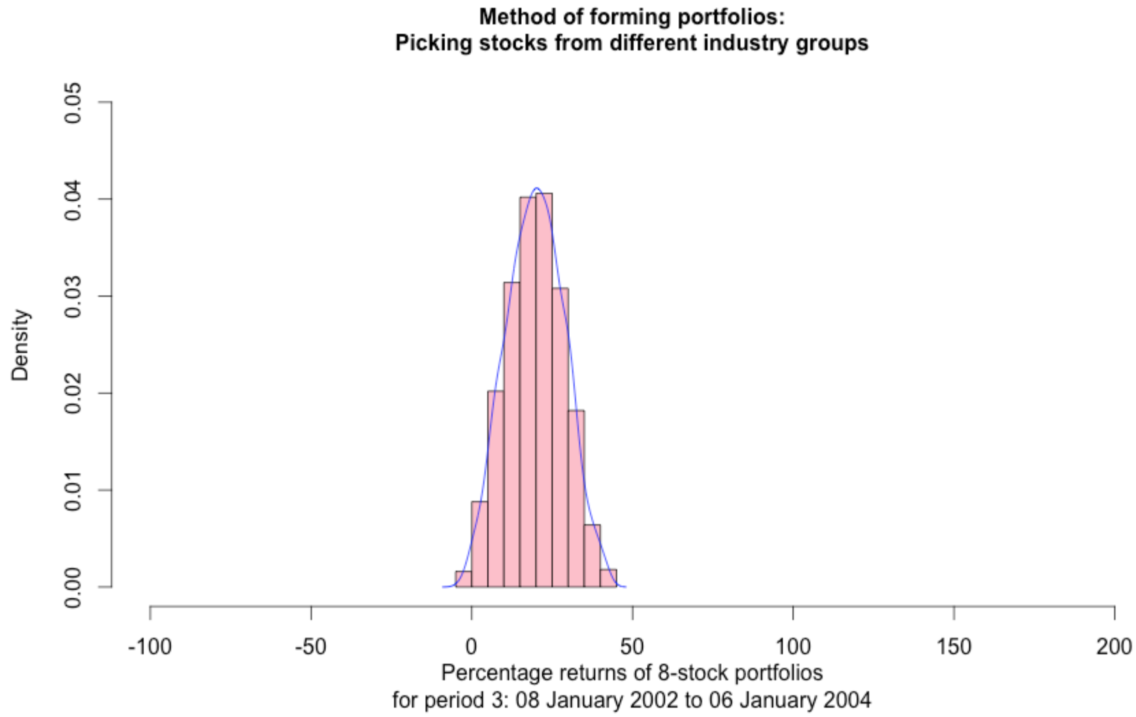


Figure 5.15 (e). The histogram and density curve of the simulated portfolios' percentage returns for period 3: 08 January 2002 to 06 January 2004. Each portfolio contains eight stocks which were picked from different industry groups.

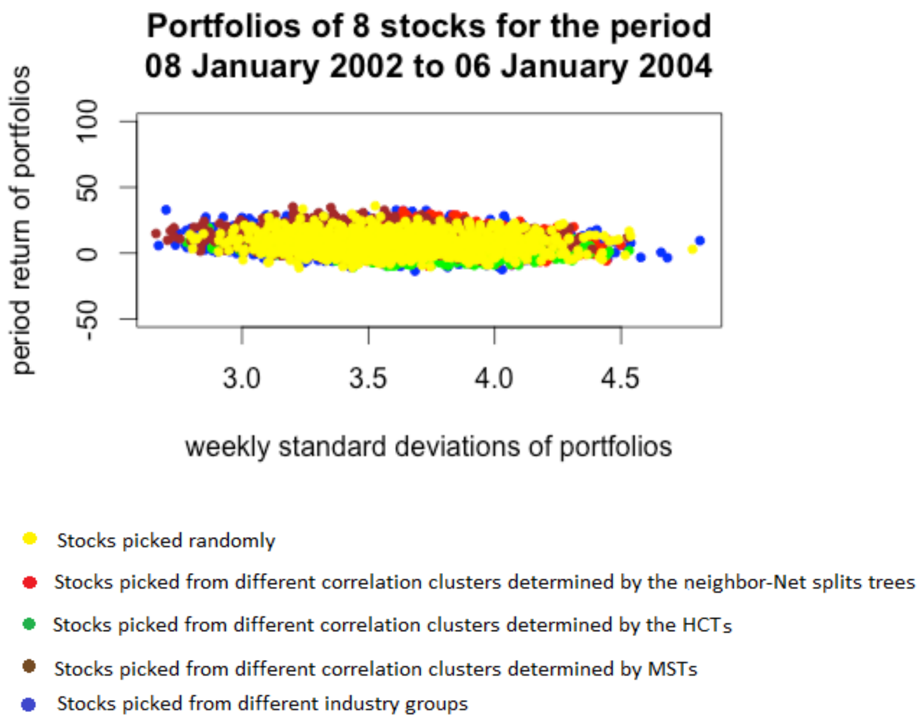


Figure 5.15 (f). Weekly standard deviation and return of each set of simulated 8-stock portfolios for period 3 using the five methods. The period returns and weekly standard deviations are expressed in percentage.

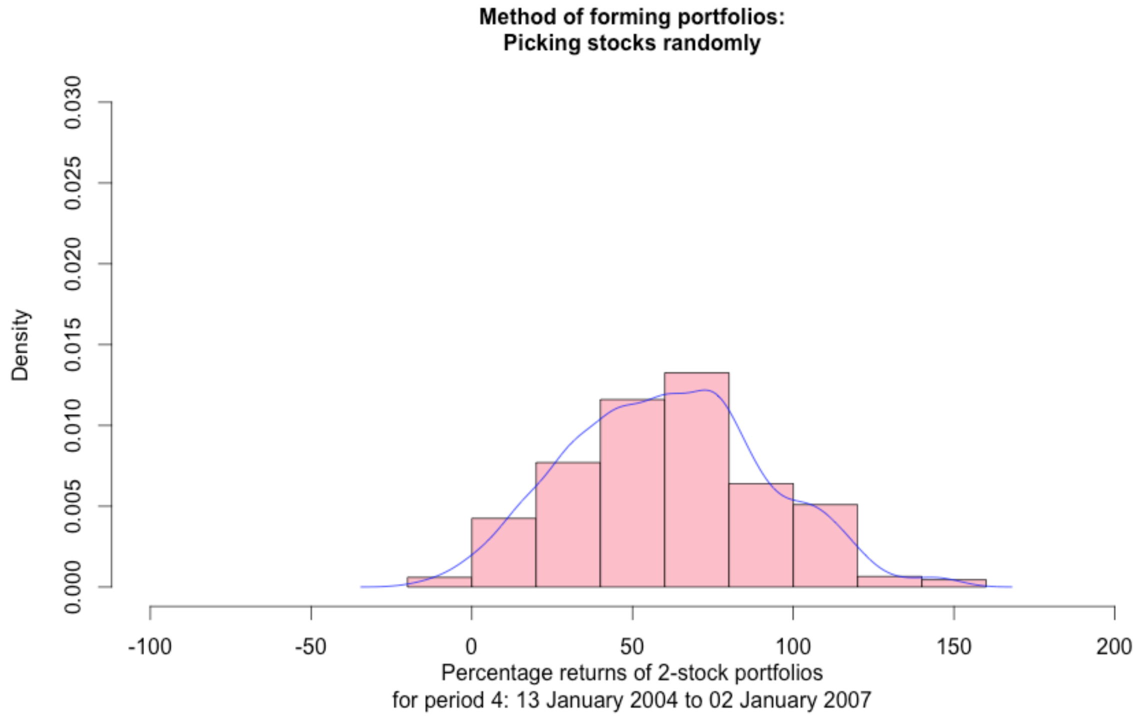


Figure 5.16 (a). The histogram and density curve of the simulated portfolios' percentage returns for period 4: 13 January 2004 to 02 January 2007. Each portfolio contains two stocks which were picked randomly.

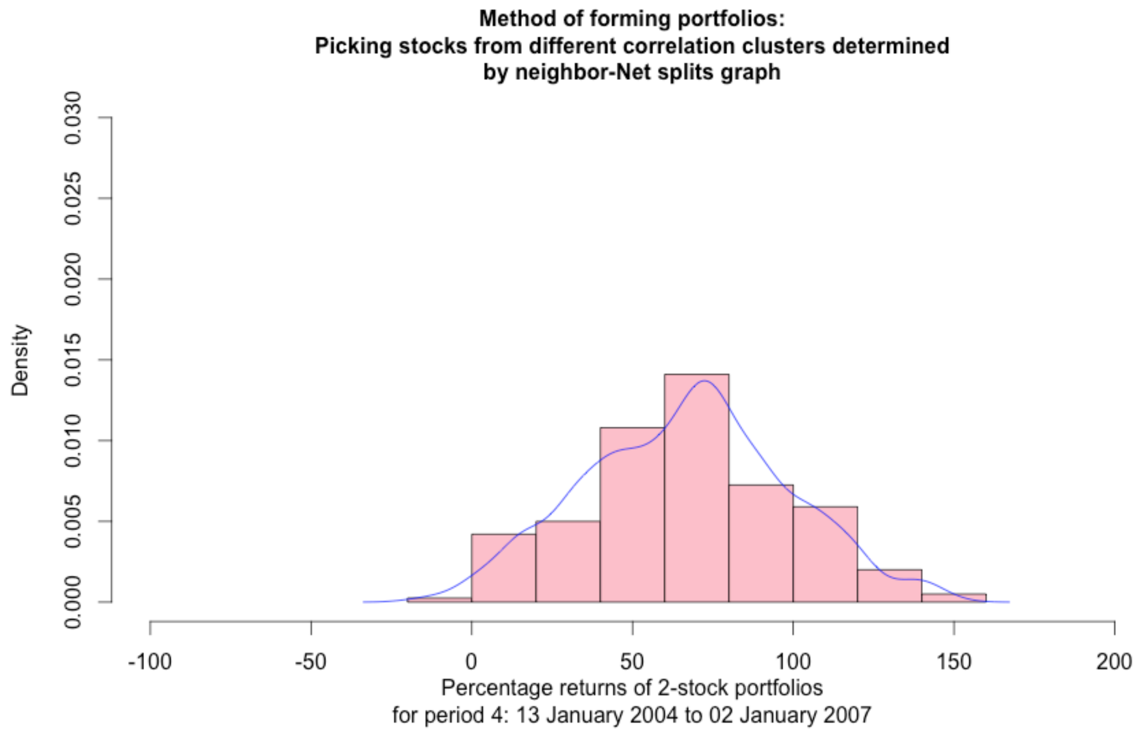


Figure 5.16 (b). The histogram and density curve of the simulated portfolios' percentage returns for period 4: 13 January 2004 to 02 January 2007. Each portfolio contains two stocks which were picked from different correlation clusters suggested by the neighbor-Net split graph produced from the stocks weekly returns in period 3: 08 January 2002 to 06 January 2004.

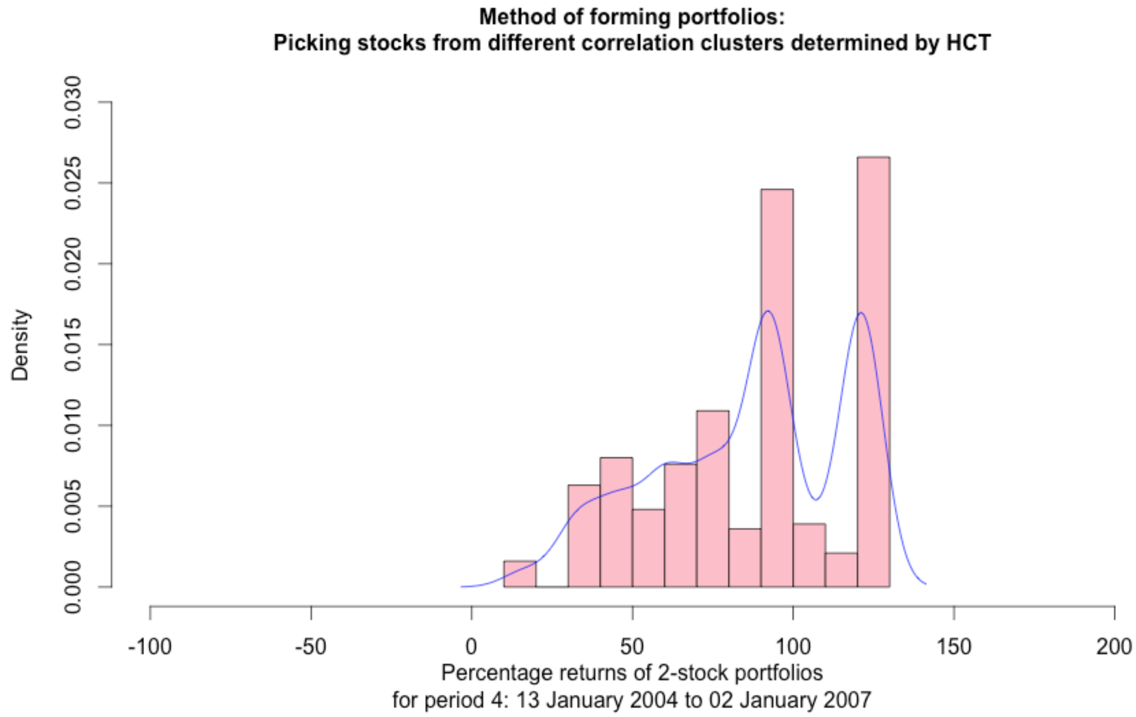


Figure 5.16 (c). The histogram and density curve of the simulated portfolios' percentage returns for period 4: 13 January 2004 to 02 January 2007. Each portfolio contains two stocks which were picked from different correlation clusters suggested by the HCT produced from the stocks weekly returns in period 3: 08 January 2002 to 06 January 2004.

Figure 5.16 (d). The histogram and density curve of the simulated portfolios' percentage returns for period 4: 13 January 2004 to 02 January 2007. Each portfolio contains two stocks which were picked from different correlation clusters suggested by the MST produced from the stocks weekly returns in period 3: 08 January 2002 to 06 January 2004.

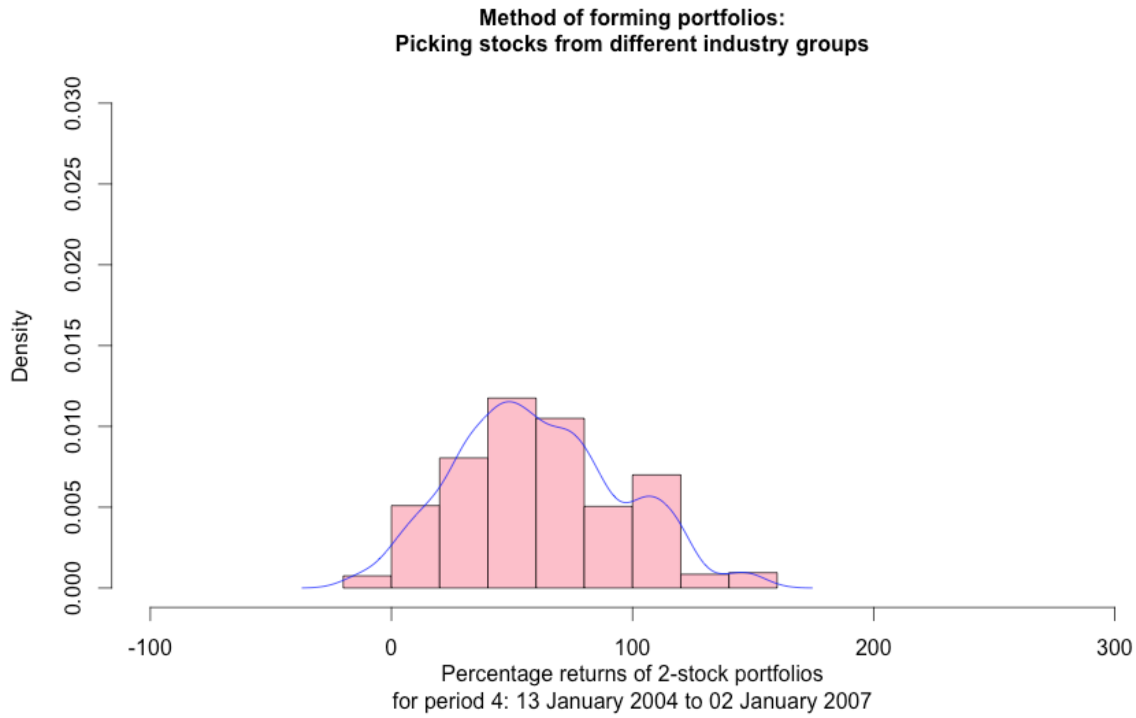


Figure 5.16 (e). The histogram and density curve of the simulated portfolios' percentage returns for period 4: 13 January 2004 to 02 January 2007. Each portfolio contains two stocks which were picked from different industry group.

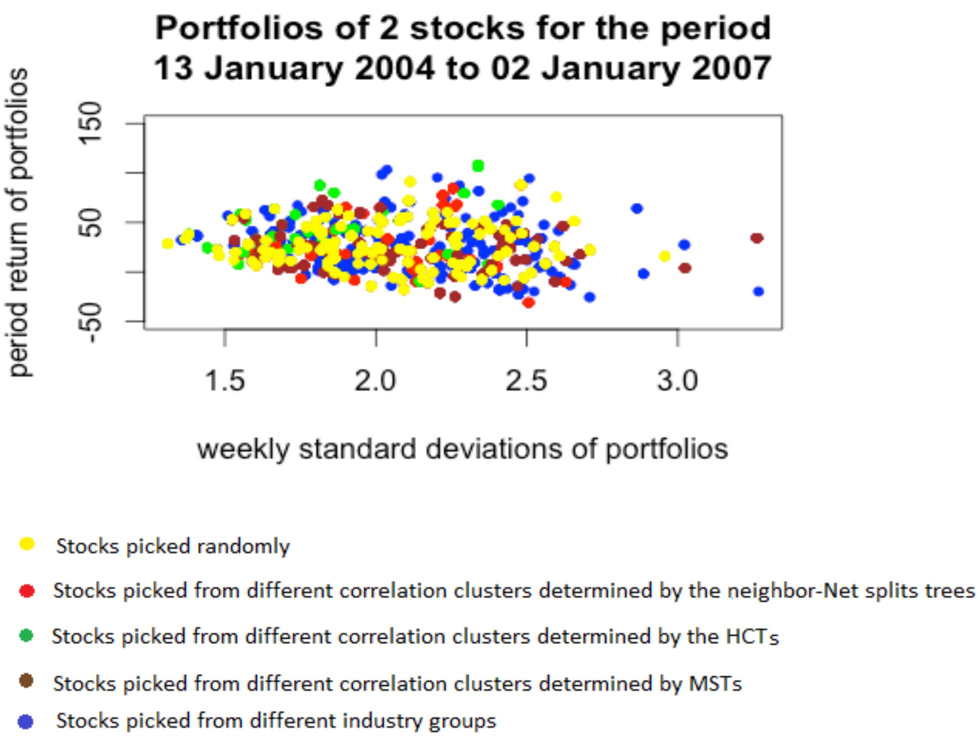


Figure 5.16 (f) Weekly standard deviation and return of each set of simulated 2-stock portfolios for period 4 using the five methods. The period returns and weekly standard deviations are expressed in percentage.

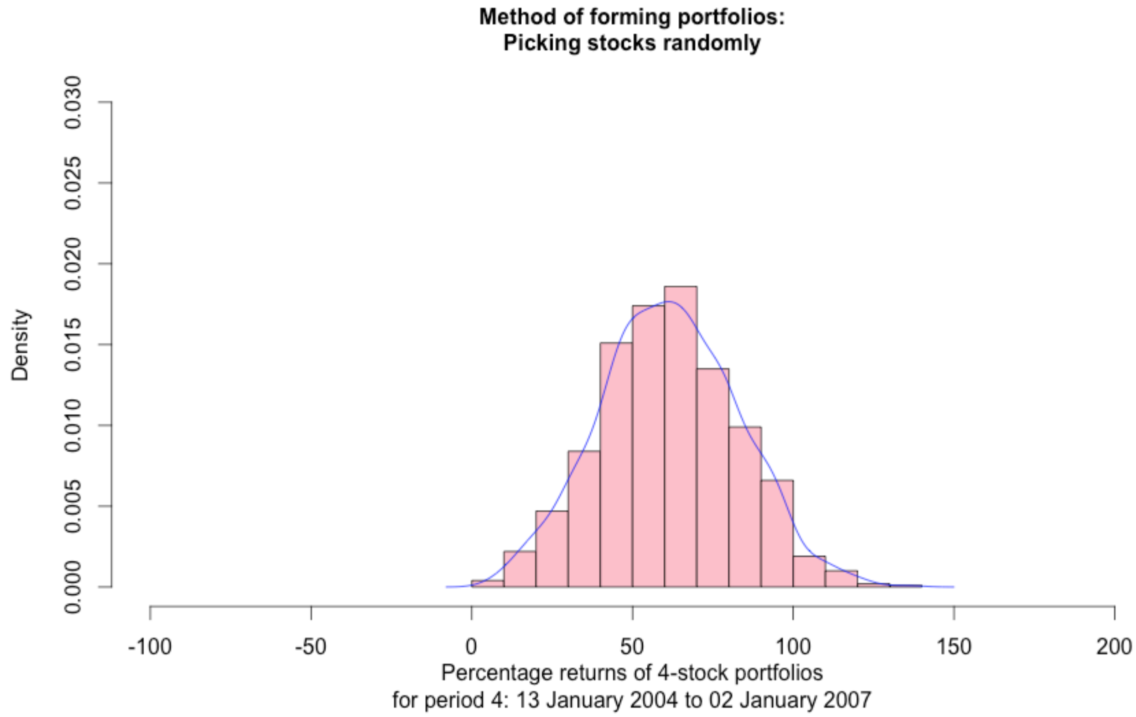


Figure 5.17 (a). The histogram and density curve of the simulated portfolios' percentage returns for period 4: 13 January 2004 to 02 January 2007. Each portfolio contains four stocks which were picked randomly.

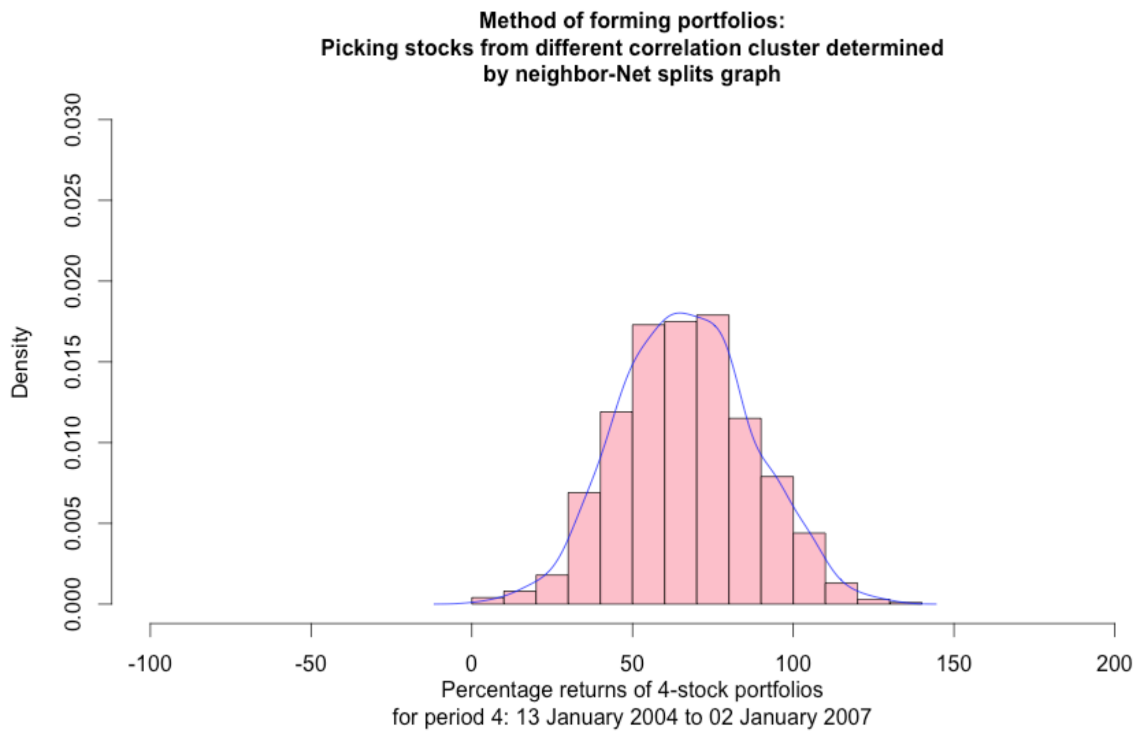


Figure 5.17 (b). The histogram and density curve of the simulated portfolios' percentage returns for period 4:13 January 2004 to 02 January 2007. Each portfolio contains four stocks which were picked from different correlation clusters suggested by the neighbor-Net split graph produced from the stocks weekly returns in period 3: 08 January 2002 to 06 January 2004.

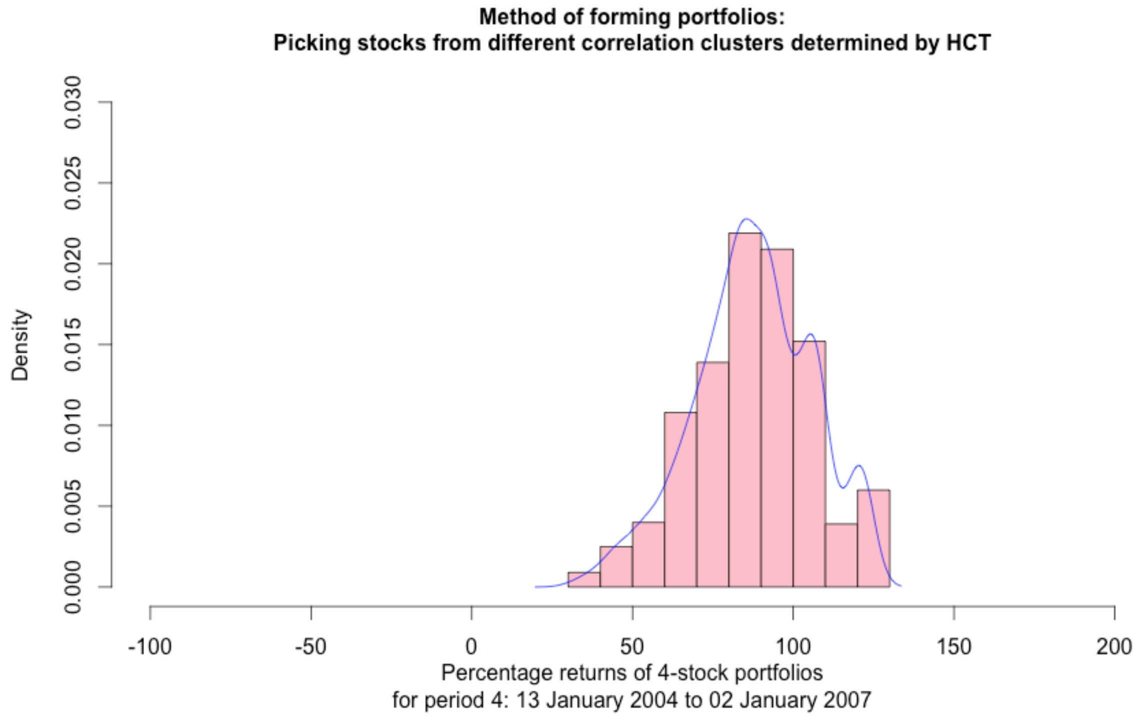


Figure 5.17 (c). The histogram and density curve of the simulated portfolios' percentage returns for period 4: 13 January 2004 to 02 January 2007. Each portfolio contains four stocks which were picked from different correlation clusters suggested by the HCT produced from the stocks weekly returns in period 3: 08 January 2002 to 06 January 2004.

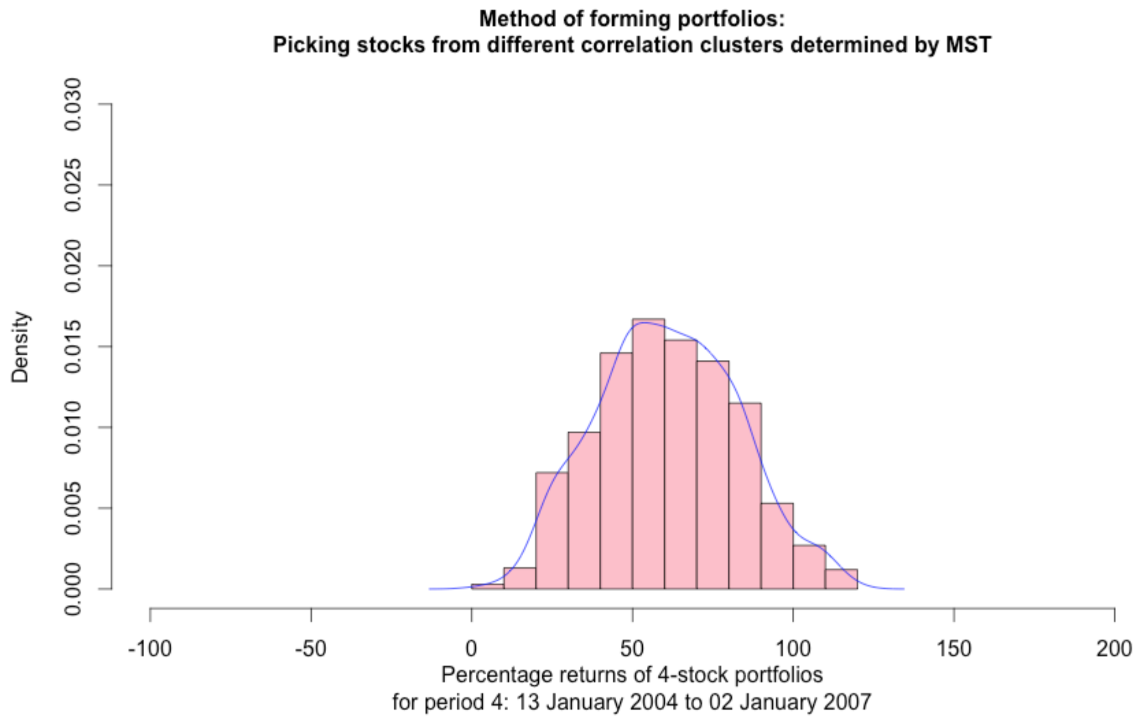


Figure 5.17 (d). The histogram and density curve of the simulated portfolios' percentage returns for period 4: 13 January 2004 to 02 January 2007. Each portfolio contains four stocks which were picked from different correlation clusters suggested by the MST produced from the stocks weekly returns in period 3: 08 January 2002 to 06 January 2004.

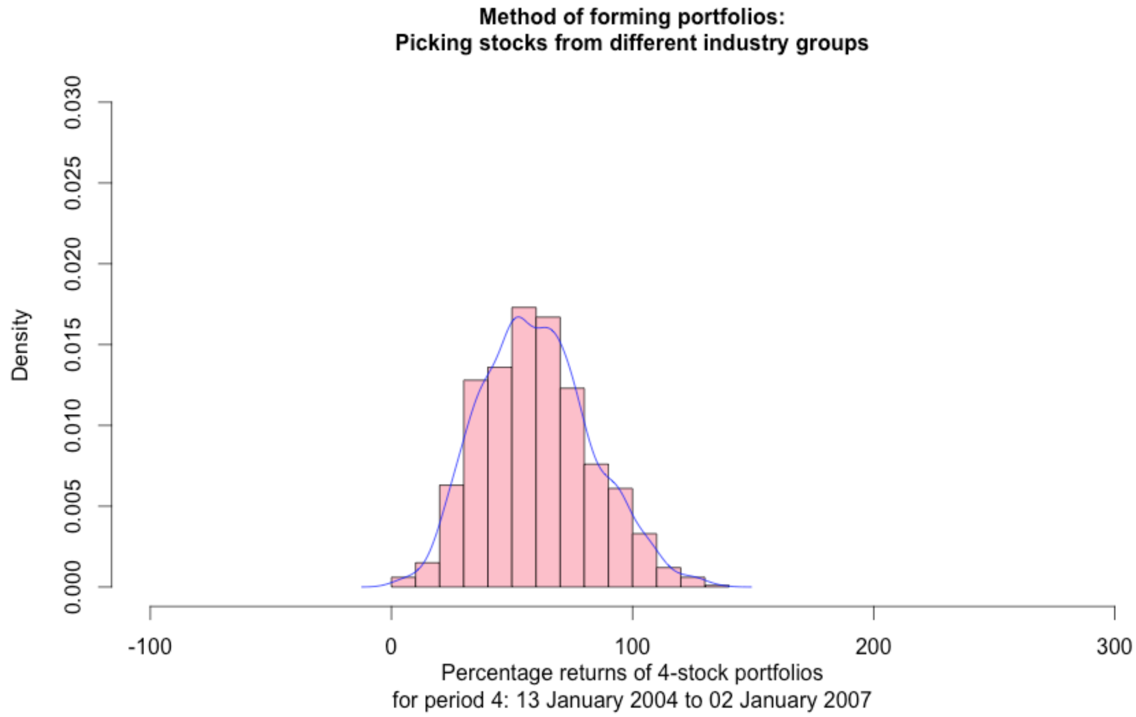


Figure 5.17 (e). The histogram and density curve of the simulated portfolios' percentage returns for period 4: 13 January 2004 to 02 January 2007. Each portfolio contains four stocks which were picked from different industry group.

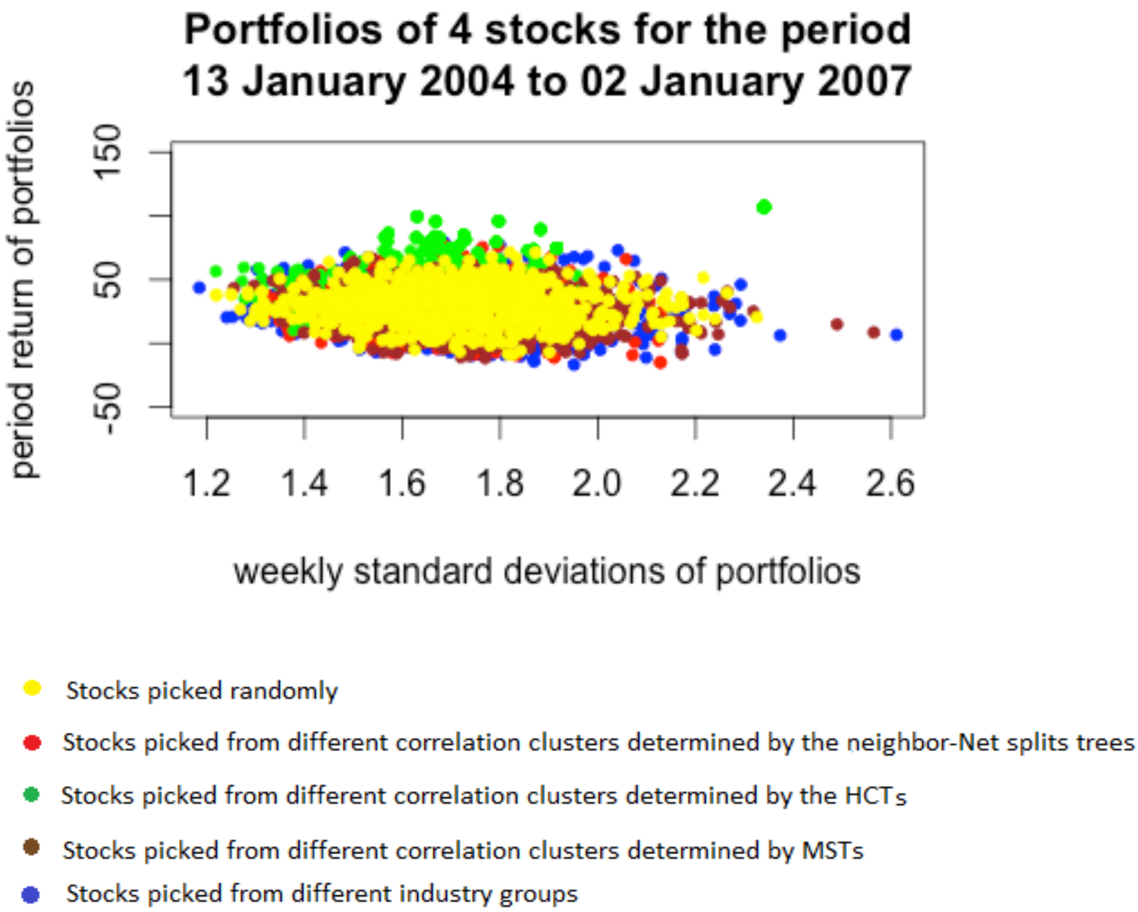


Figure 5.17 (f) Weekly standard deviation and return of each set of simulated 4-stock portfolios for period 4 using the five methods. The period returns and weekly standard deviations are expressed in percentage.

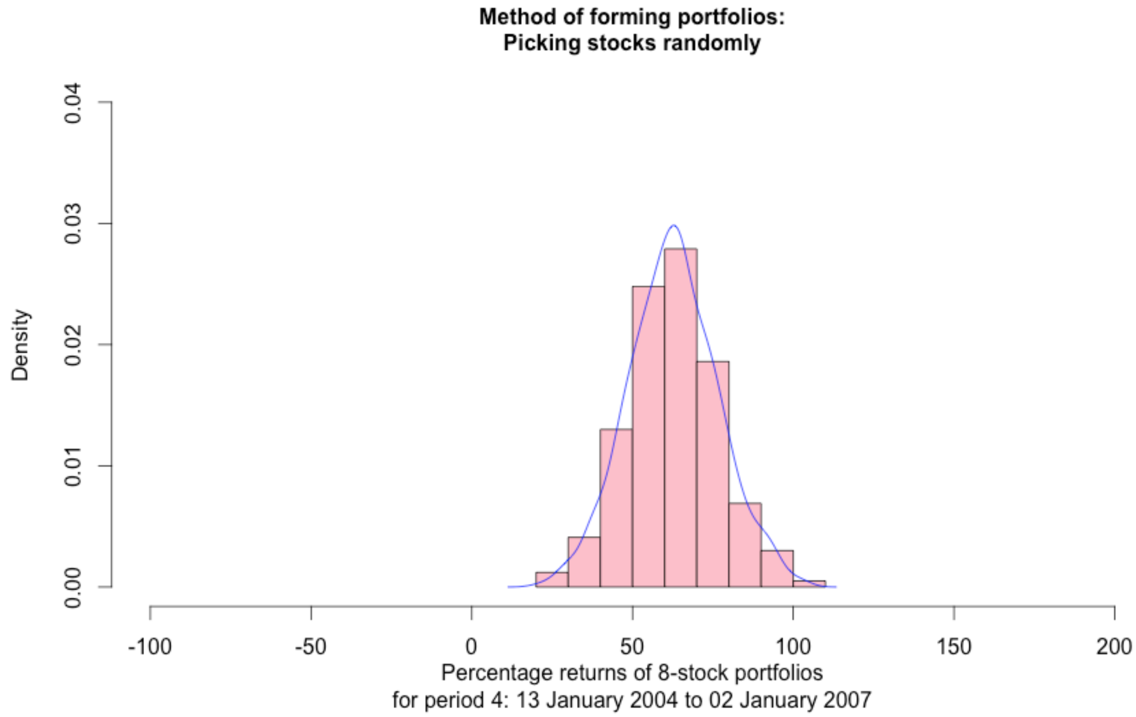


Figure 5.18 (a). The histogram and density curve of the simulated portfolios' percentage returns for period 4: 13 January 2004 to 02 January 2007. Each portfolio contains eight stocks which were picked randomly.

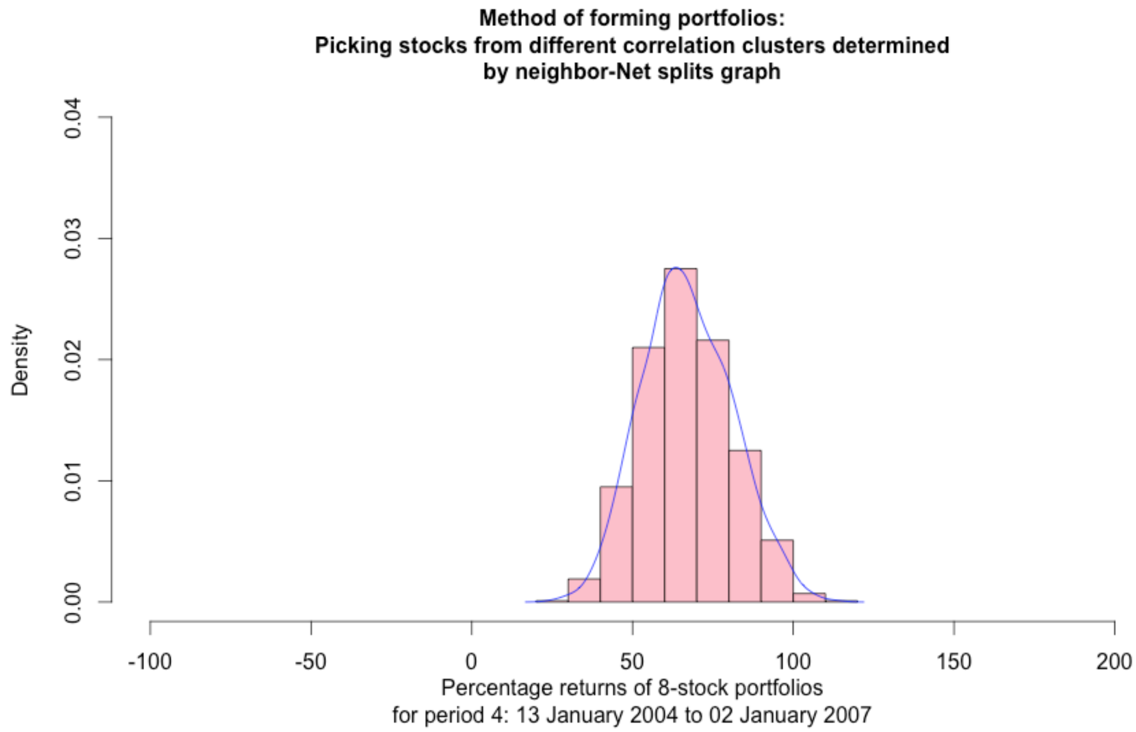


Figure 5.18 (b). The histogram and density curve of the simulated portfolios' percentage returns for period 4: 13 January 2004 to 02 January 2007. Each portfolio contains eight stocks which were picked from different correlation clusters suggested by the neighbor-Net split graph produced from the stocks weekly returns in period 3: 08 January 2002 to 06 January 2004.

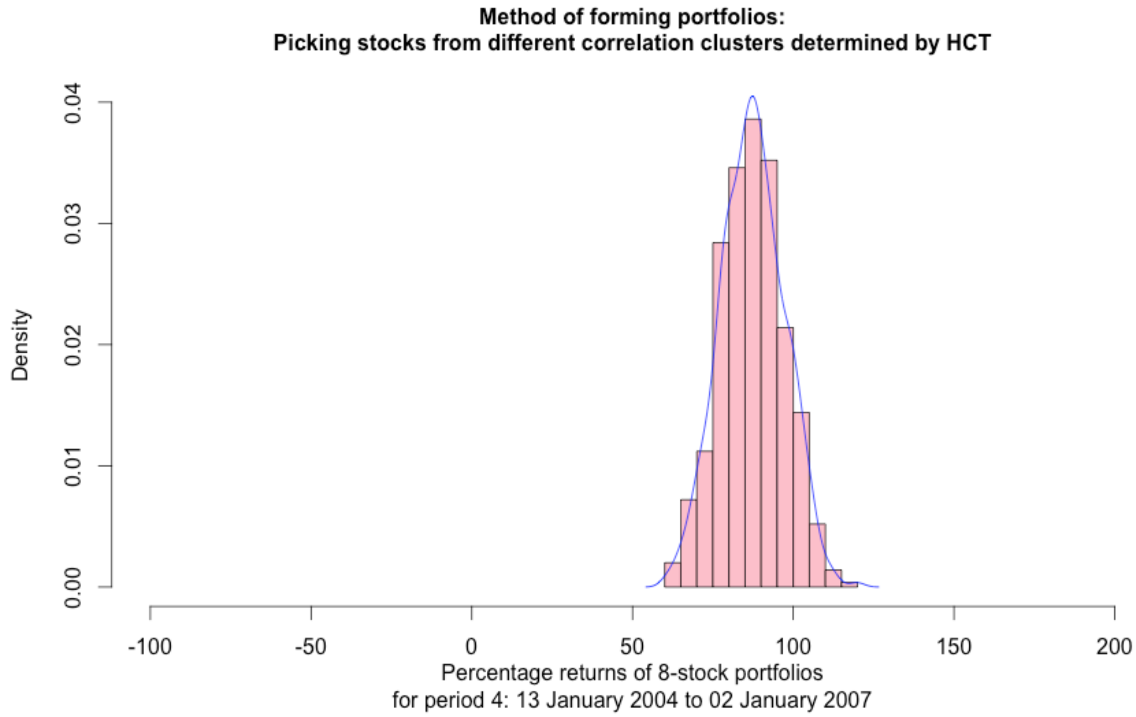


Figure 5.18 (c). The histogram and density curve of the simulated portfolios' percentage returns for period 4: 13 January 2004 to 02 January 2007. Each portfolio contains eight stocks which were picked from different correlation clusters suggested by the HCT produced from the stocks weekly returns in period 3: 08 January 2002 to 06 January 2004.

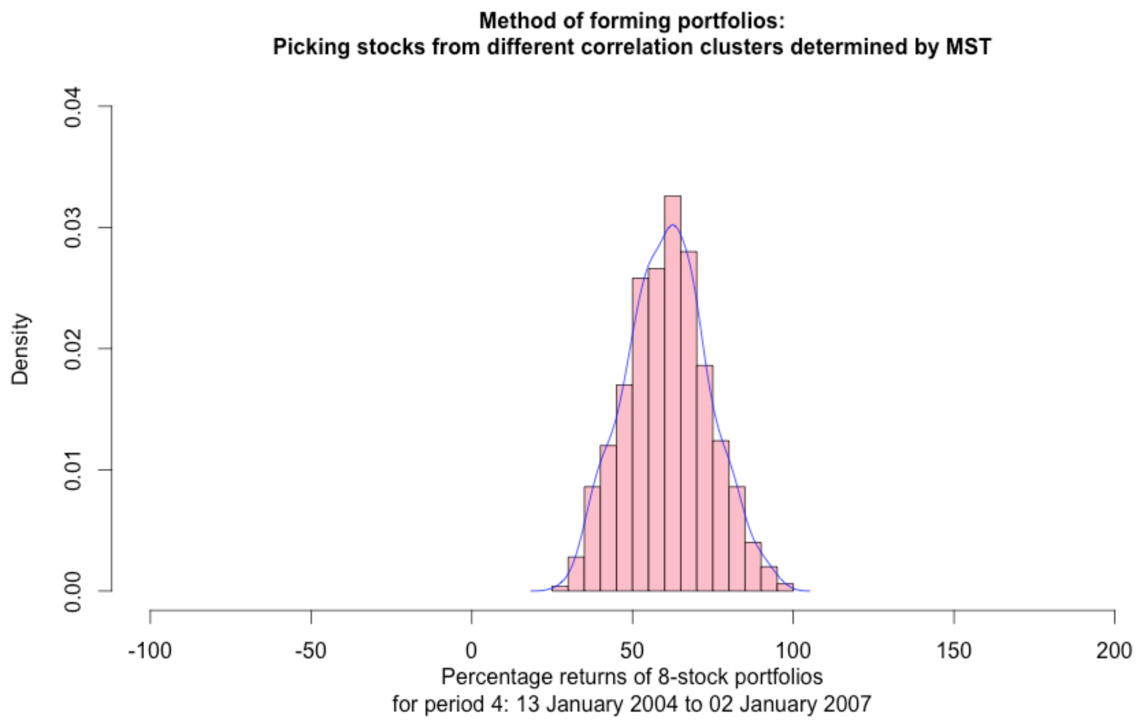


Figure 5.18 (d). The histogram and density curve of the simulated portfolios' percentage returns for period 4: 13 January 2004 to 02 January 2007. Each portfolio contains eight stocks which were picked from different correlation clusters suggested by the MST produced from the stocks weekly returns in period 3: 08 January 2002 to 06 January 2004.

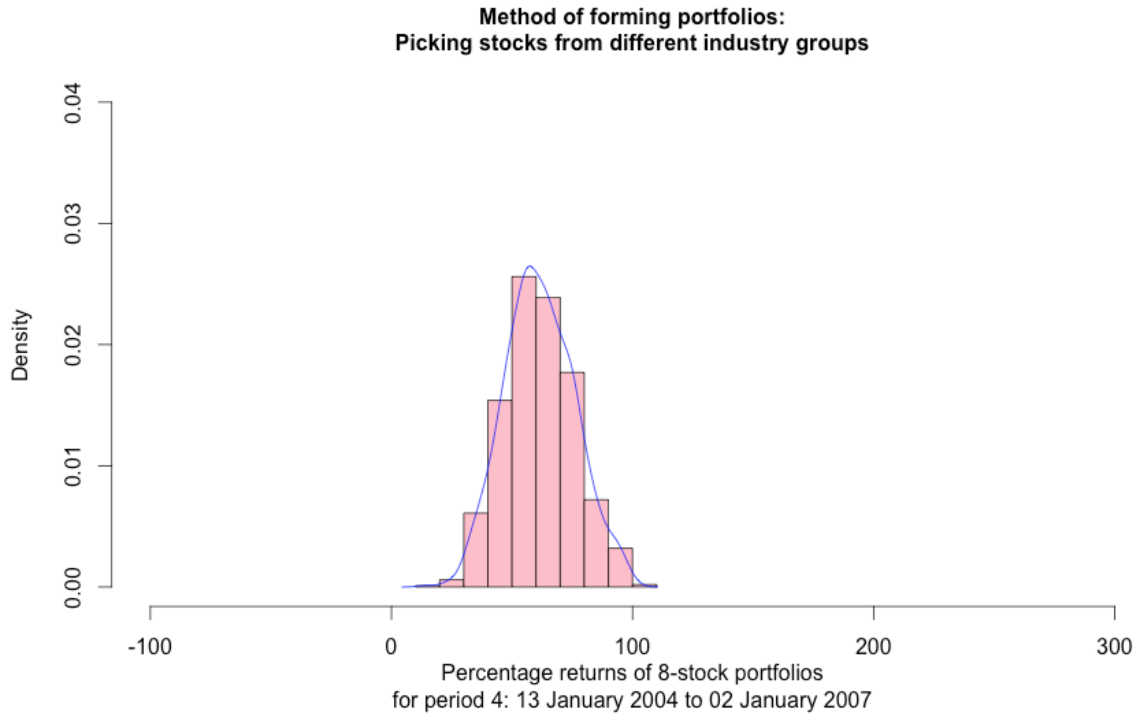


Figure 5.18 (e). The histogram and density curve of the simulated portfolios' percentage returns for period 4: 13 January 2004 to 02 January 2007. Each portfolio contains eight stocks which were picked from different industry group.

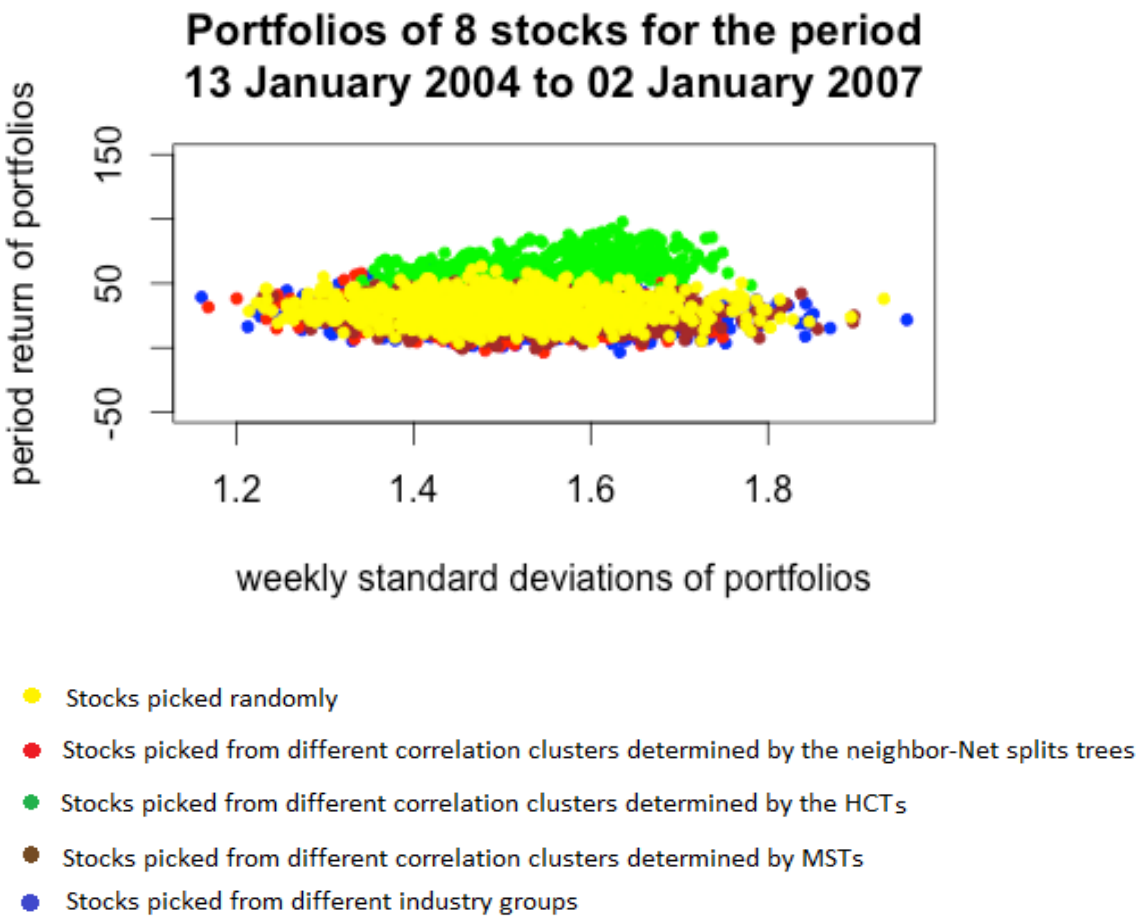


Figure 5.18 (f) Weekly standard deviation and return of each set of simulated 8-stock portfolios for period 4 using the five methods. The period returns and weekly standard deviations are expressed in percentage.

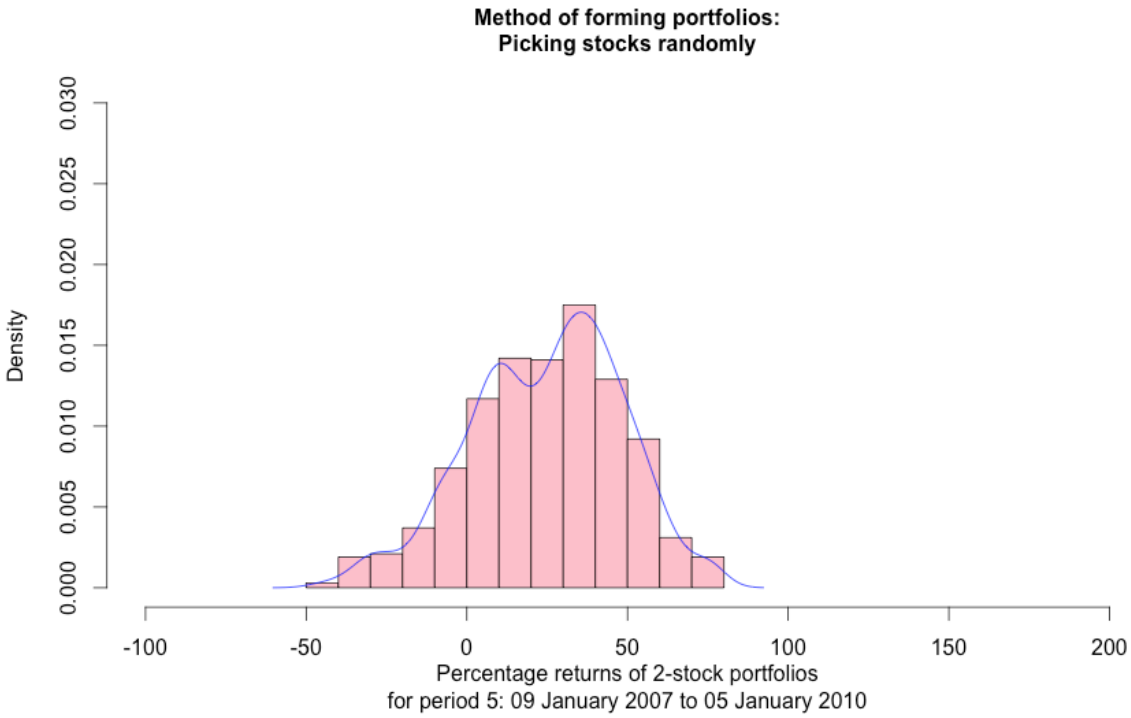


Figure 5.19 (a). The histogram and density curve of the simulated portfolios' percentage returns for period 5: 09 January 2007 to 05 January 2010. Each portfolio contains two stocks which were picked randomly.

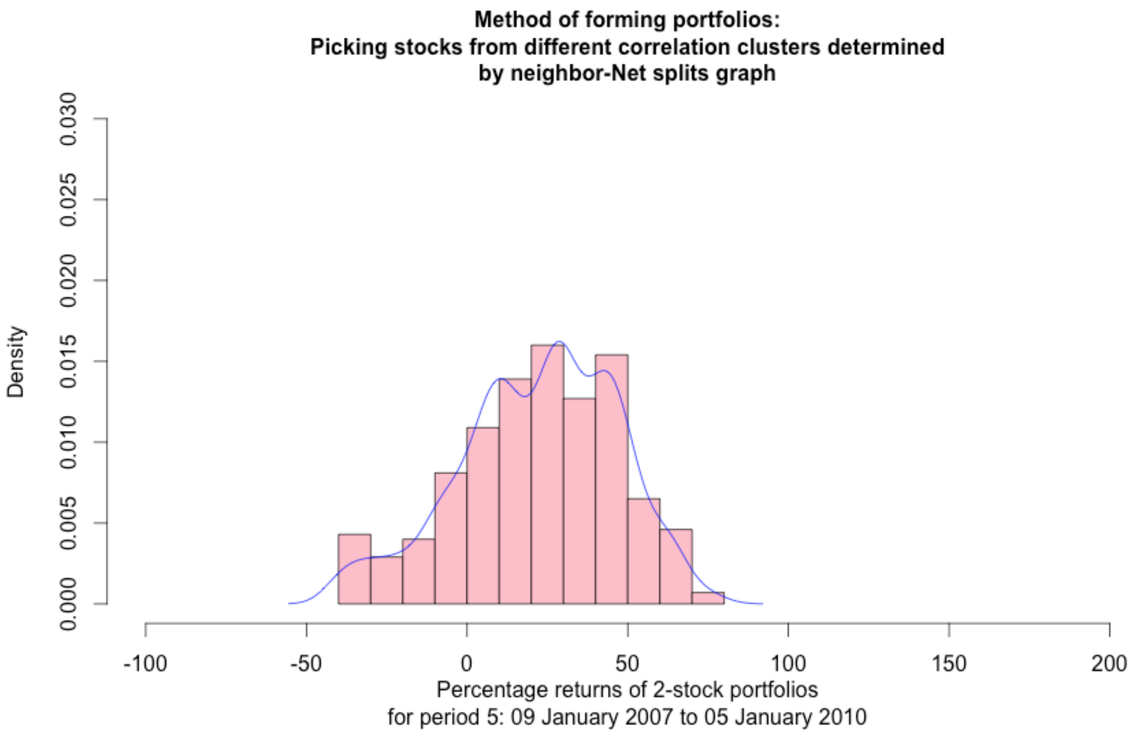


Figure 5.19 (b). The histogram and density curve of the simulated portfolios' percentage returns for period 5: 09 January 2007 to 05

January 2010. Each portfolio contains two stocks which were picked from different correlation clusters suggested by the neighbor-Net split graph produced from the stocks weekly returns in period 4: 13 January 2004 to 02 January 2007.

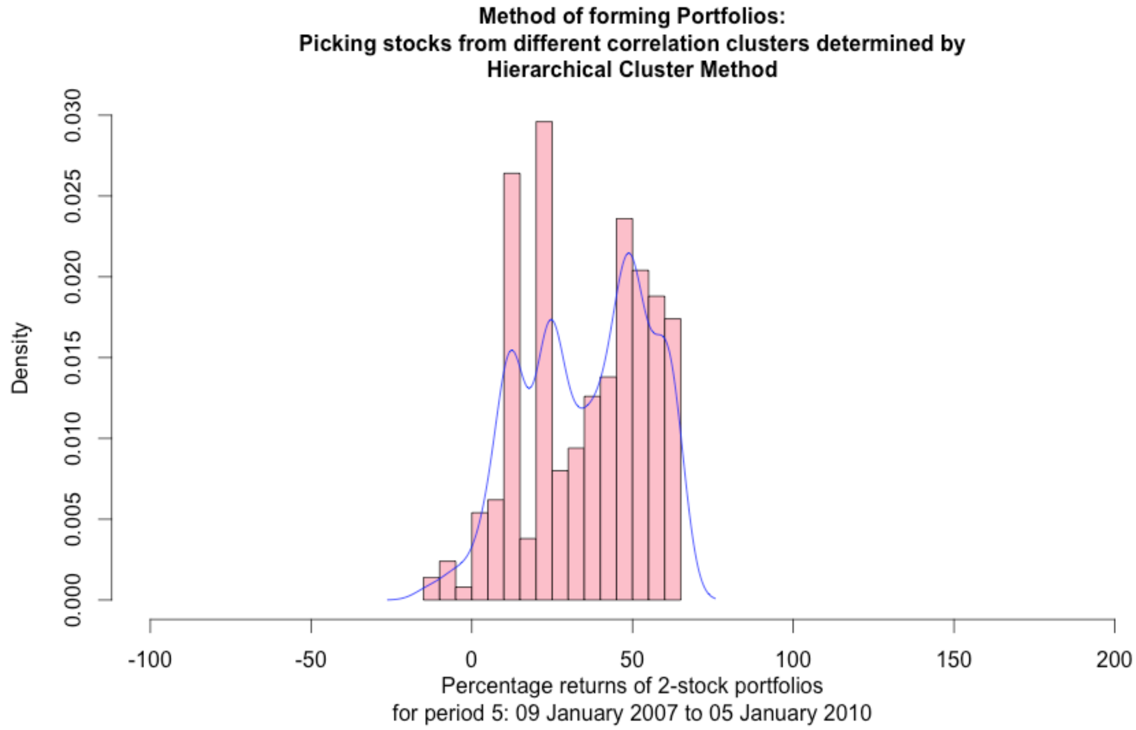


Figure 5.19 (c). The histogram and density curve of the simulated portfolios' percentage returns for period 5: 09 January 2007 to 05 January 2010. Each portfolio contains two stocks which were picked from different correlation clusters suggested by the HCT produced from the stocks weekly returns in period 4: 13 January 2004 to 02 January 2007.

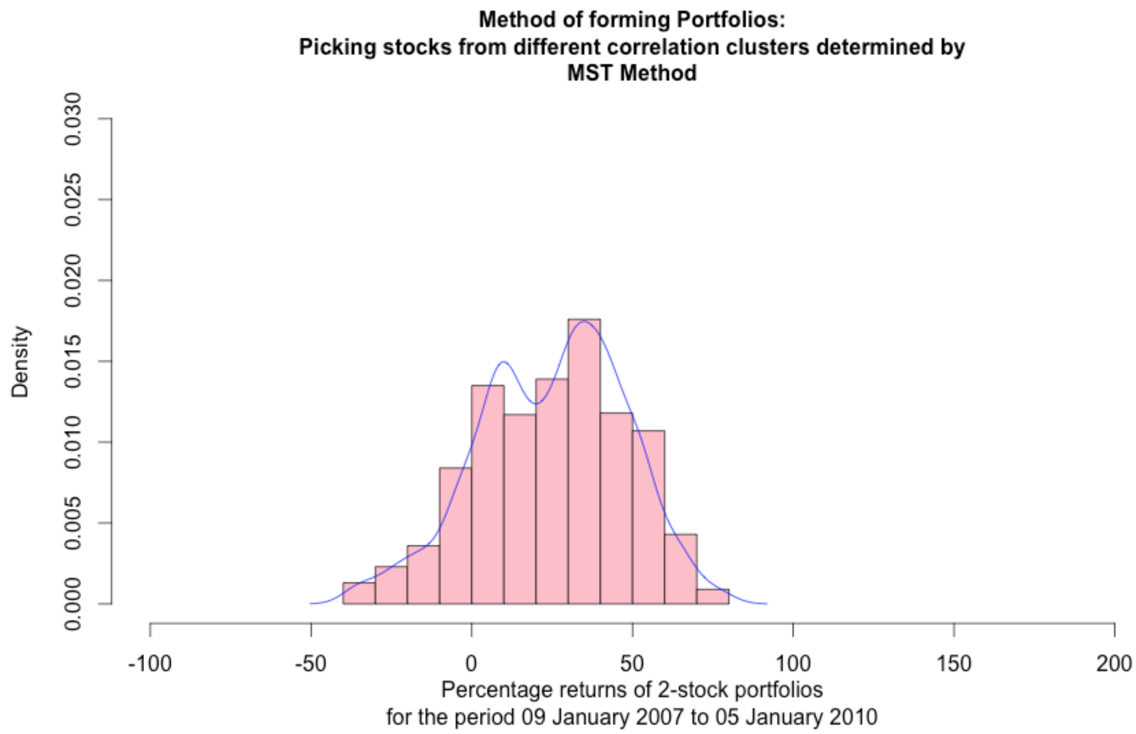


Figure 5.19 (d). The histogram and density curve of the simulated portfolios' percentage returns for period 5: 09 January 2007 to 05

January 2010. Each portfolio contains two stocks which were picked from different correlation clusters suggested by the MST produced from the stocks weekly returns in period 4: 13 January 2004 to 02 January 2007.

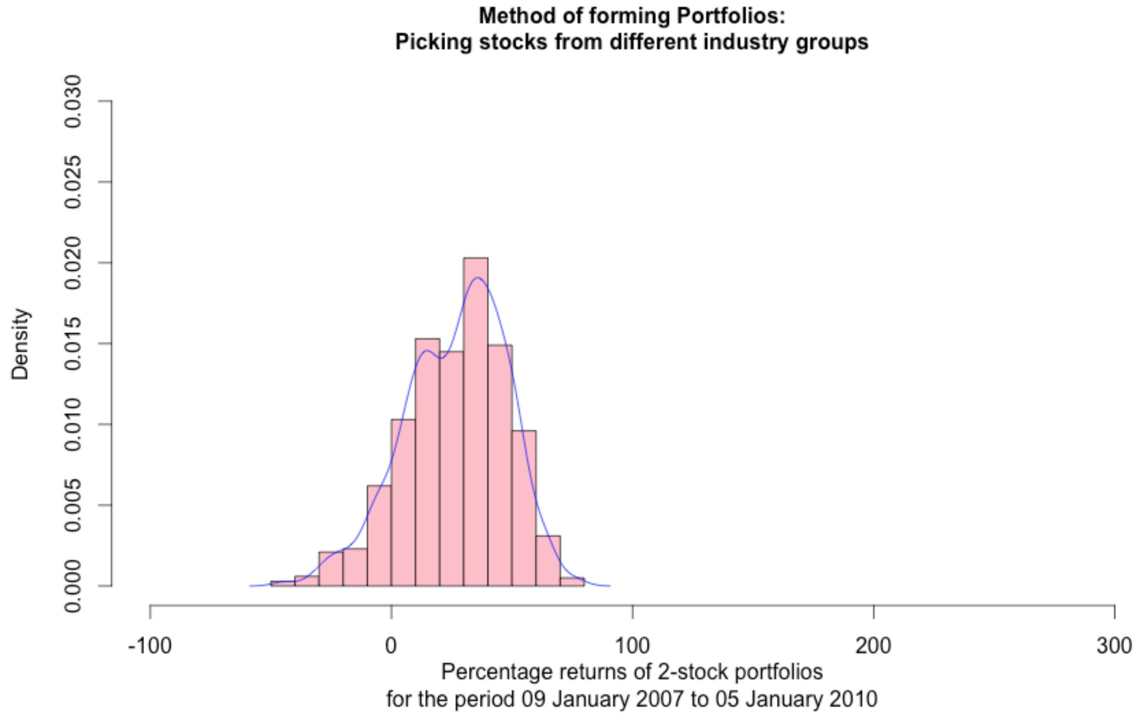


Figure 5.19 (e). The histogram and density curve of the simulated portfolios' percentage returns for period 5: 09 January 2007 to 05 January 2010. Each portfolio contains two stocks which were picked from different industry groups.

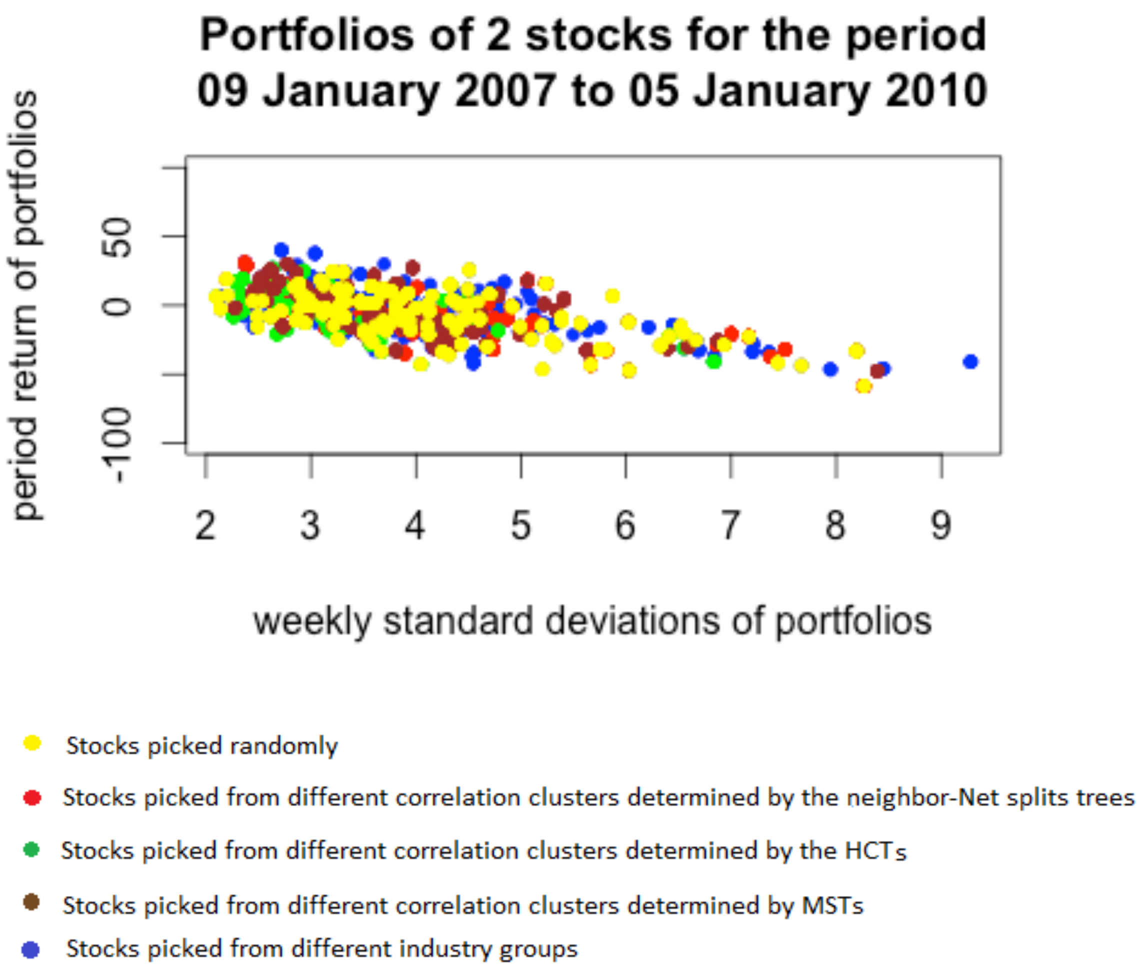


Figure 5.19 (f). Weekly standard deviation and return of each set of simulated 2-stock portfolios for period 5 using the five methods. The period returns and weekly standard deviations are expressed in percentage.

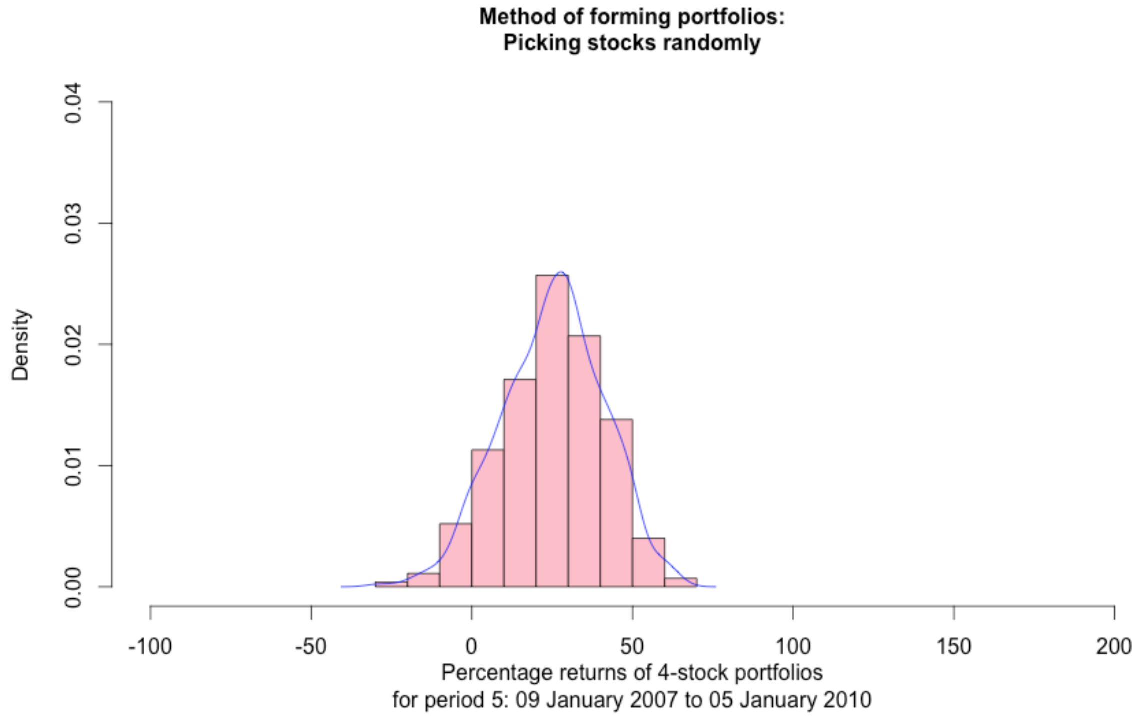


Figure 5.20 (a). The histogram and density curve of the simulated portfolios' percentage returns for period 5: 09 January 2007 to 05 January 2010. Each portfolio contains four stocks which were picked randomly.

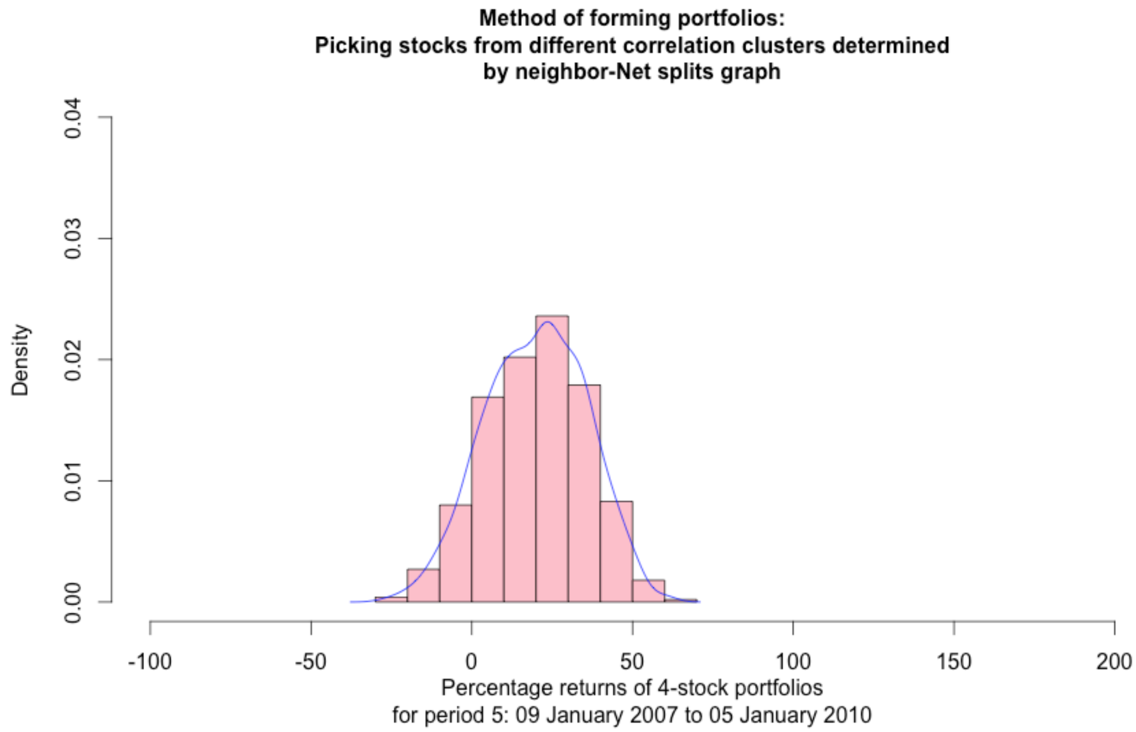


Figure 5.20 (b). The histogram and density curve of the simulated portfolios' percentage returns for period 5: 09 January 2007 to 05 January 2010. Each portfolio contains four stocks which were picked from different correlation clusters suggested by the neighbor-Net split graph produced from the stocks weekly returns in period 4: 13 January 2004 to 02 January 2007.

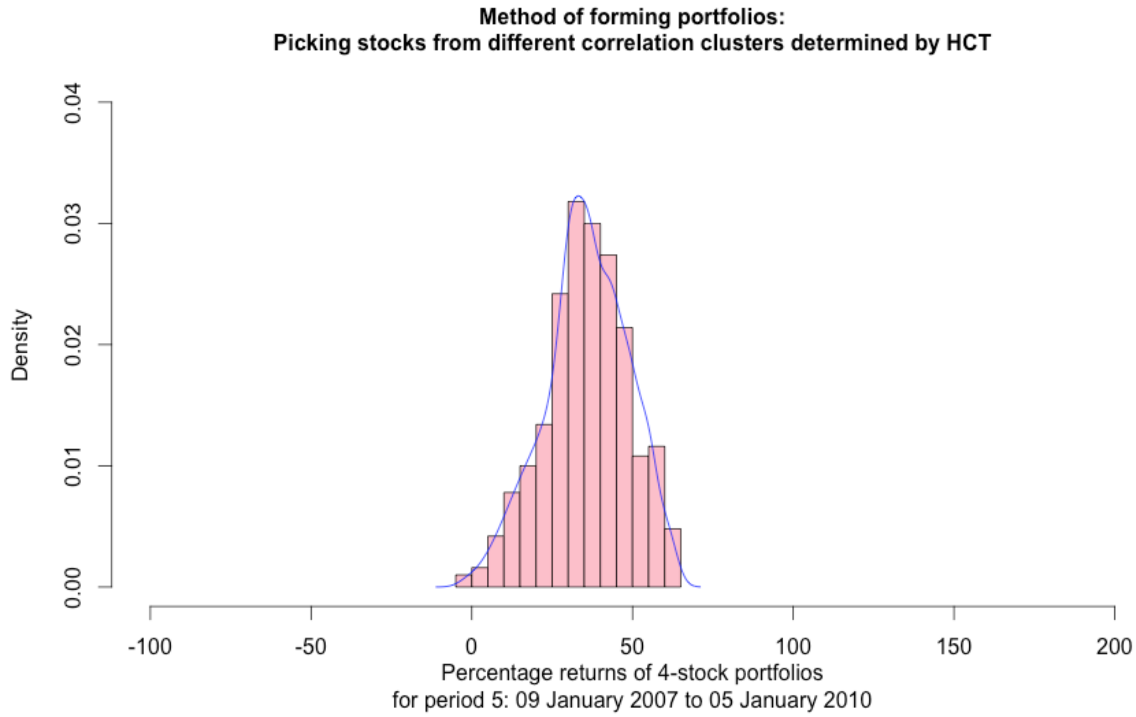


Figure 5.20 (c). The histogram and density curve of the simulated portfolios' percentage returns for period 5: 09 January 2007 to 05 January 2010. Each portfolio contains four stocks which were picked from different correlation clusters suggested by the HCT produced from the stocks weekly returns in period 4:13 January 2004 to 02 January 2007.

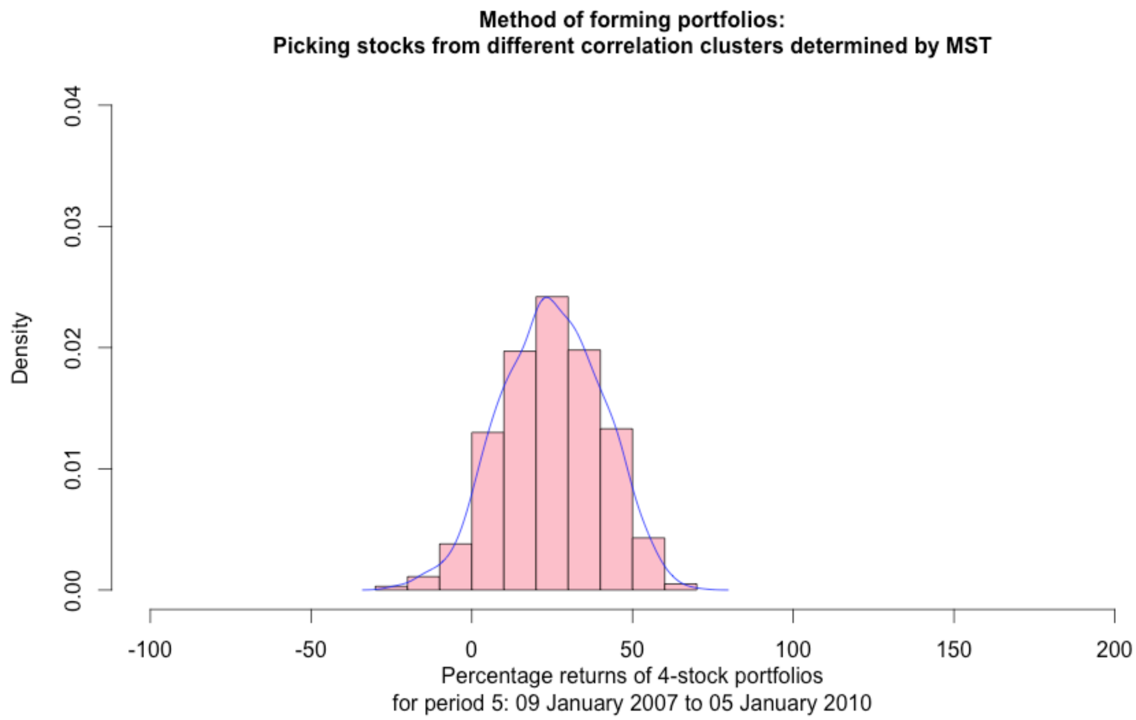


Figure 5.20 (d). The histogram and density curve of the simulated portfolios' percentage returns for period 5: 09 January 2007 to 05 January 2010. Each portfolio contains four stocks which were picked from different correlation clusters suggested by the MST produced from the stocks weekly returns in period 4: 13 January 2004 to 02 January 2007.

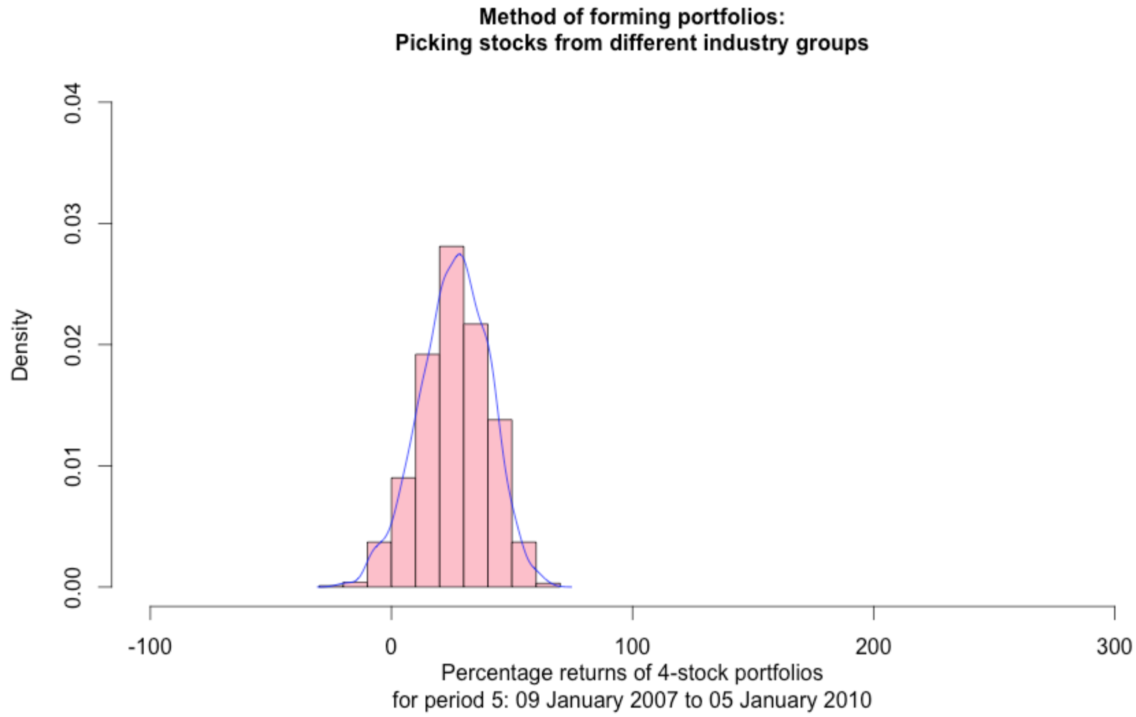


Figure 5.20 (e). The histogram and density curve of the simulated portfolios' percentage returns for period 5: 09 January 2007 to 05 January 2010. Each portfolio contains four stocks which were picked from different industry groups.

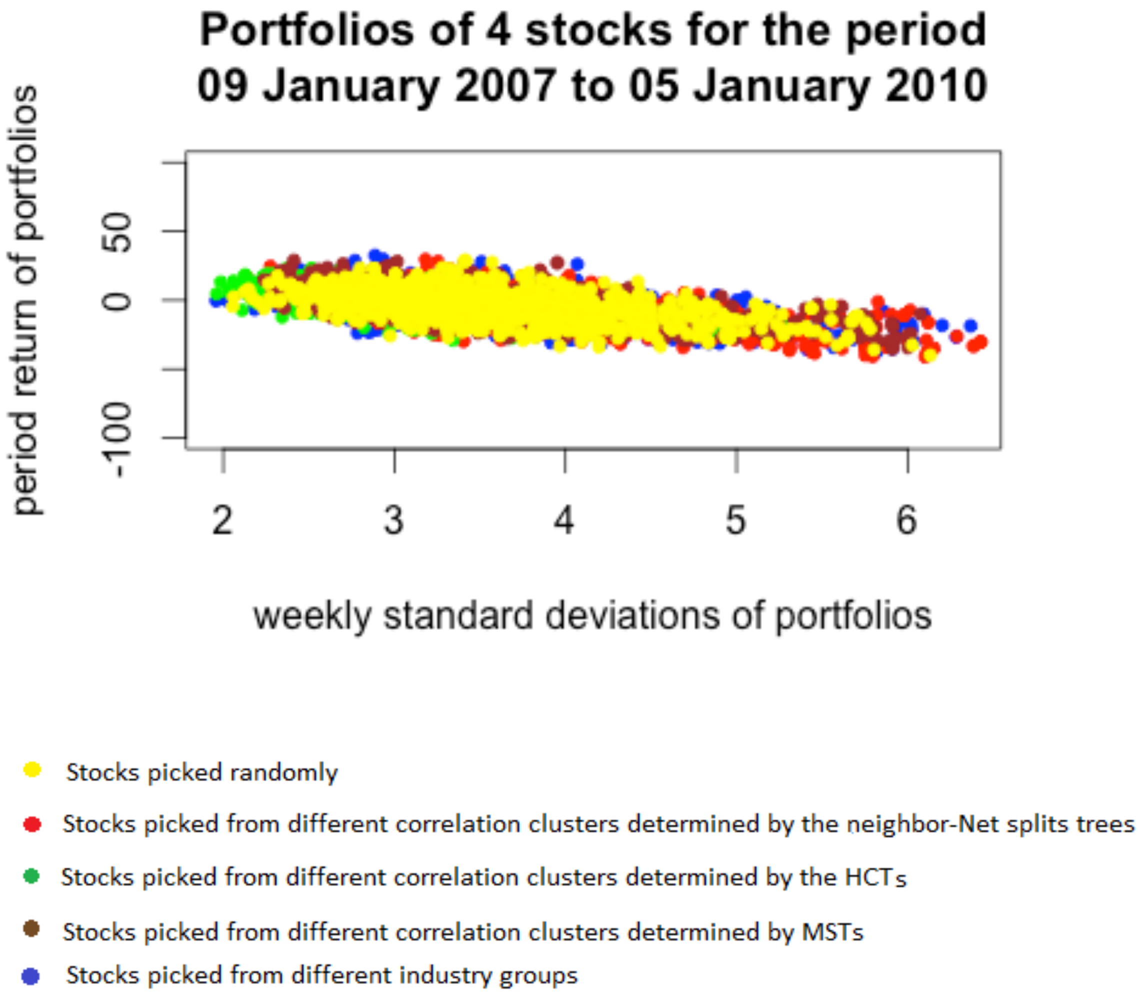


Figure 5.20 (f) Weekly standard deviation and return of each set of simulated 4-stock portfolios for period 5 using the five methods. The period returns and weekly standard deviations are expressed in percentage.

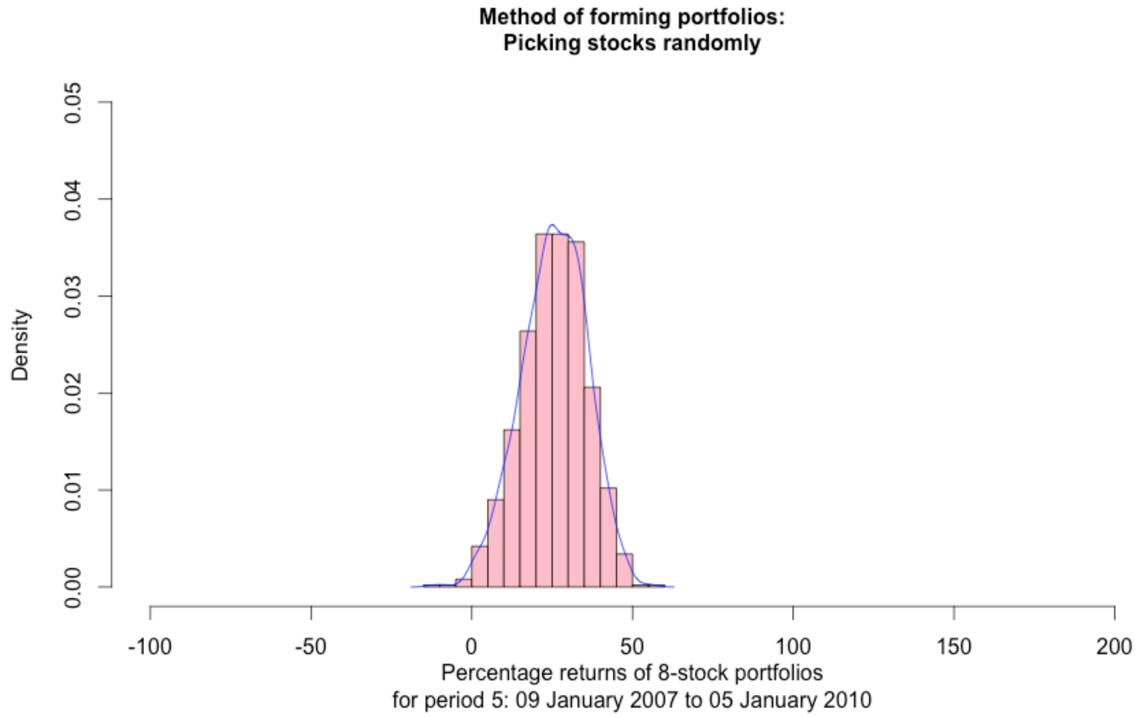


Figure 5.21 (a). The histogram and density curve of the simulated portfolios' percentage returns for period 5: 09 January 2007 to 05 January 2010. Each portfolio contains eight stocks which were picked randomly.

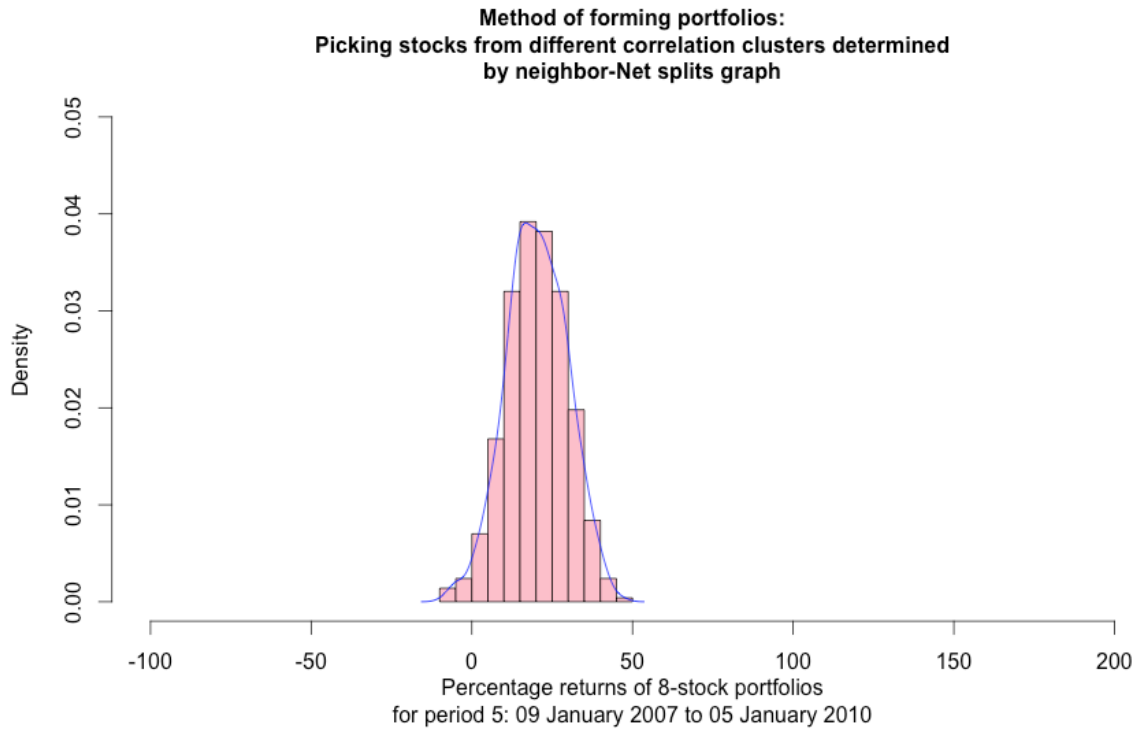


Figure 5.21 (b). The histogram and density curve of the simulated portfolios' percentage returns for period 5: 09 January 2007 to 05 January 2010. Each portfolio contains eight stocks which were picked from different correlation clusters suggested by the neighbor-Net split graph produced from the stocks weekly returns in period 4: 13 January 2004 to 02 January 2007.

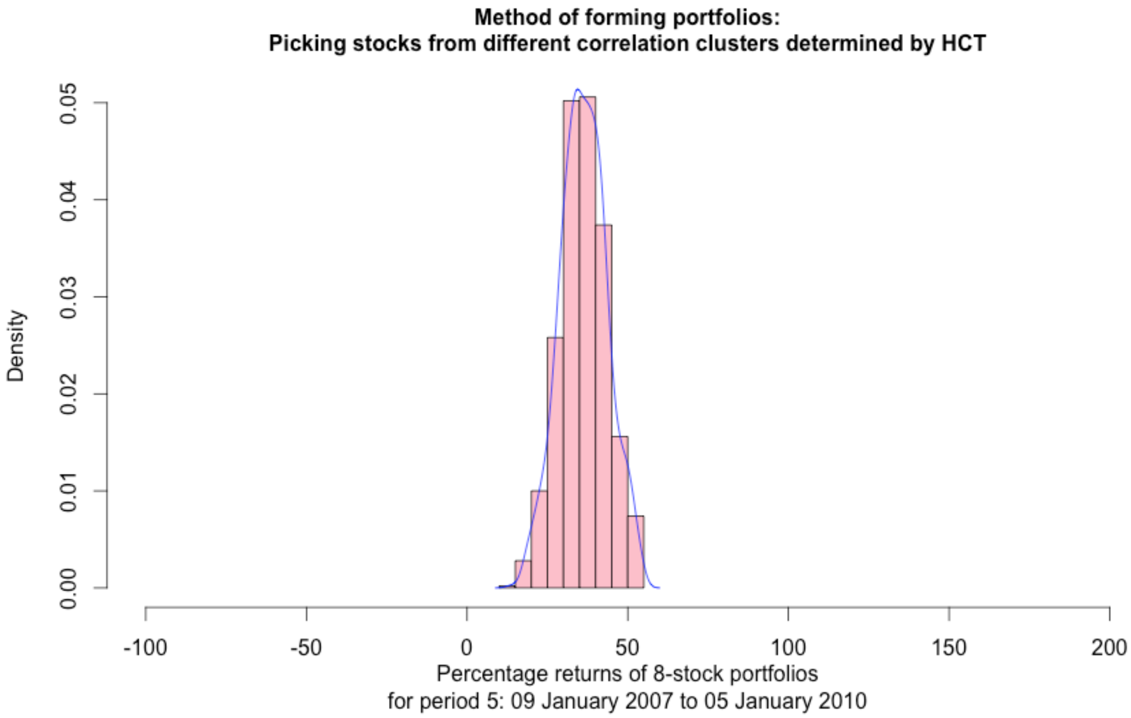


Figure 5.21 (c). The histogram and density curve of the simulated portfolios' percentage returns for period 5: 09 January 2007 to 05 January 2010. Each portfolio contains eight stocks which were picked from different correlation clusters suggested by the HCT produced from the stocks weekly returns in period 4: 13 January 2004 to 02 January 2007.

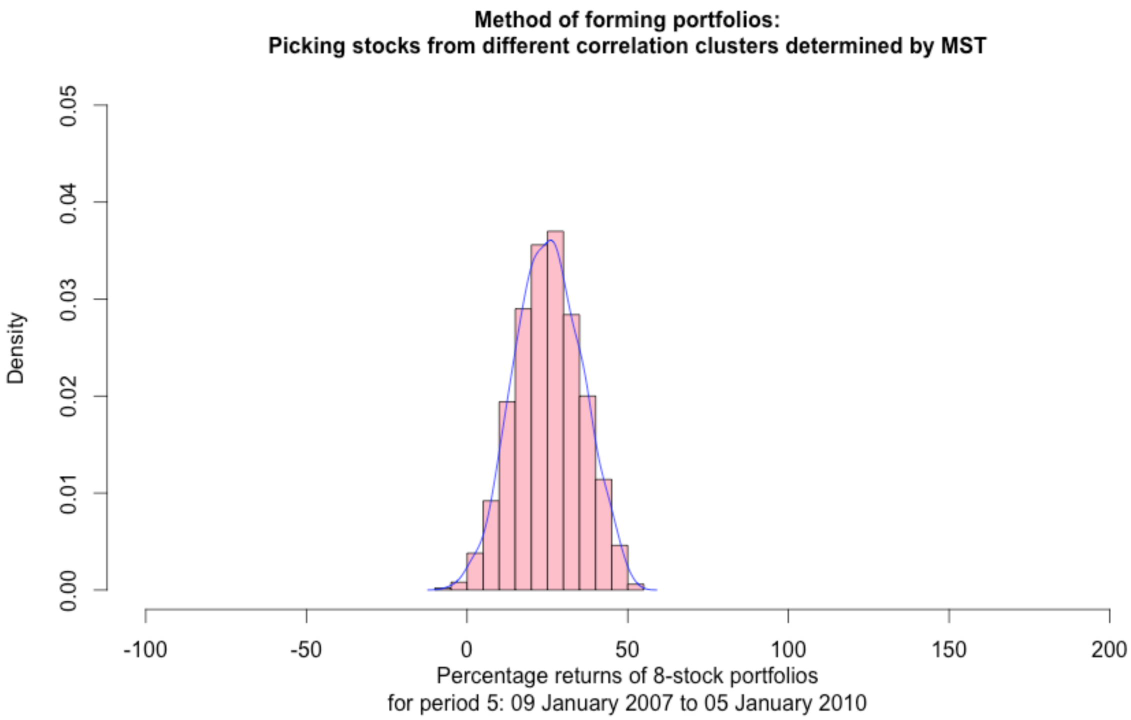


Figure 5.21 (d). The histogram and density curve of the simulated portfolios' percentage returns for period 5: 09 January 2007 to 05 January 2010. Each portfolio contains eight stocks which were picked from different correlation clusters suggested by the MST produced from the stocks weekly returns in period 4: 13 January 2004 to 02 January 2007.

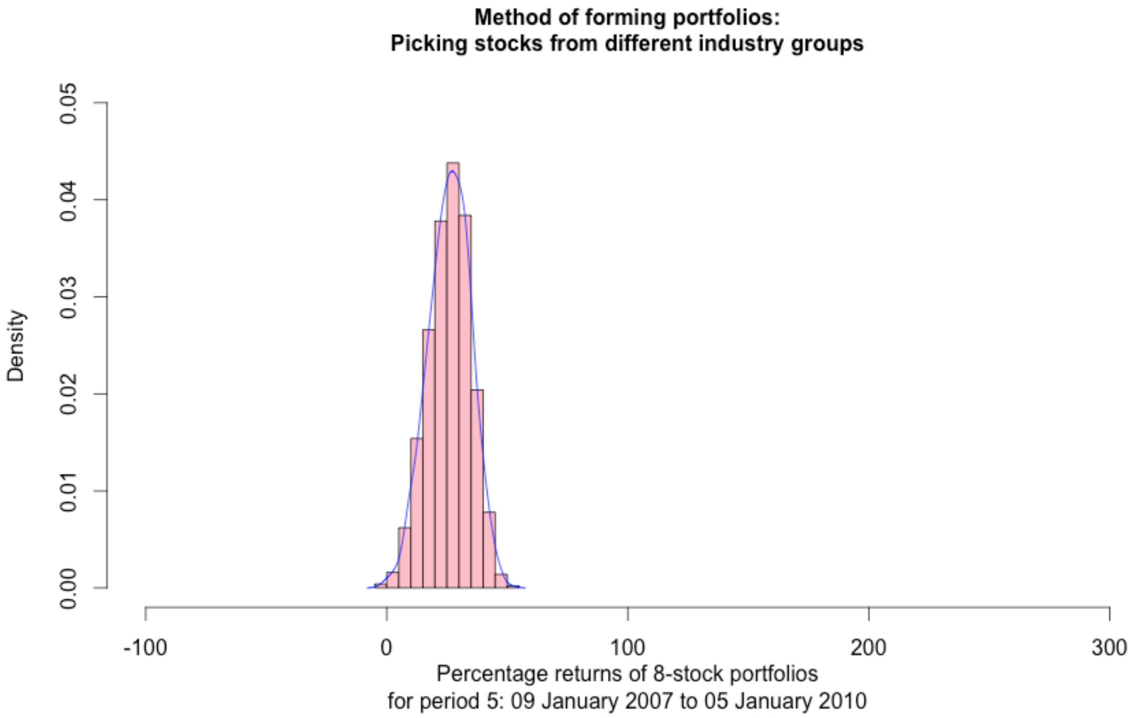


Figure 5.21 (e). The histogram and density curve of the simulated portfolios' percentage returns for period 5: 09 January 2007 to 05 January 2010. Each portfolio contains eight stocks which were picked from different industry groups.

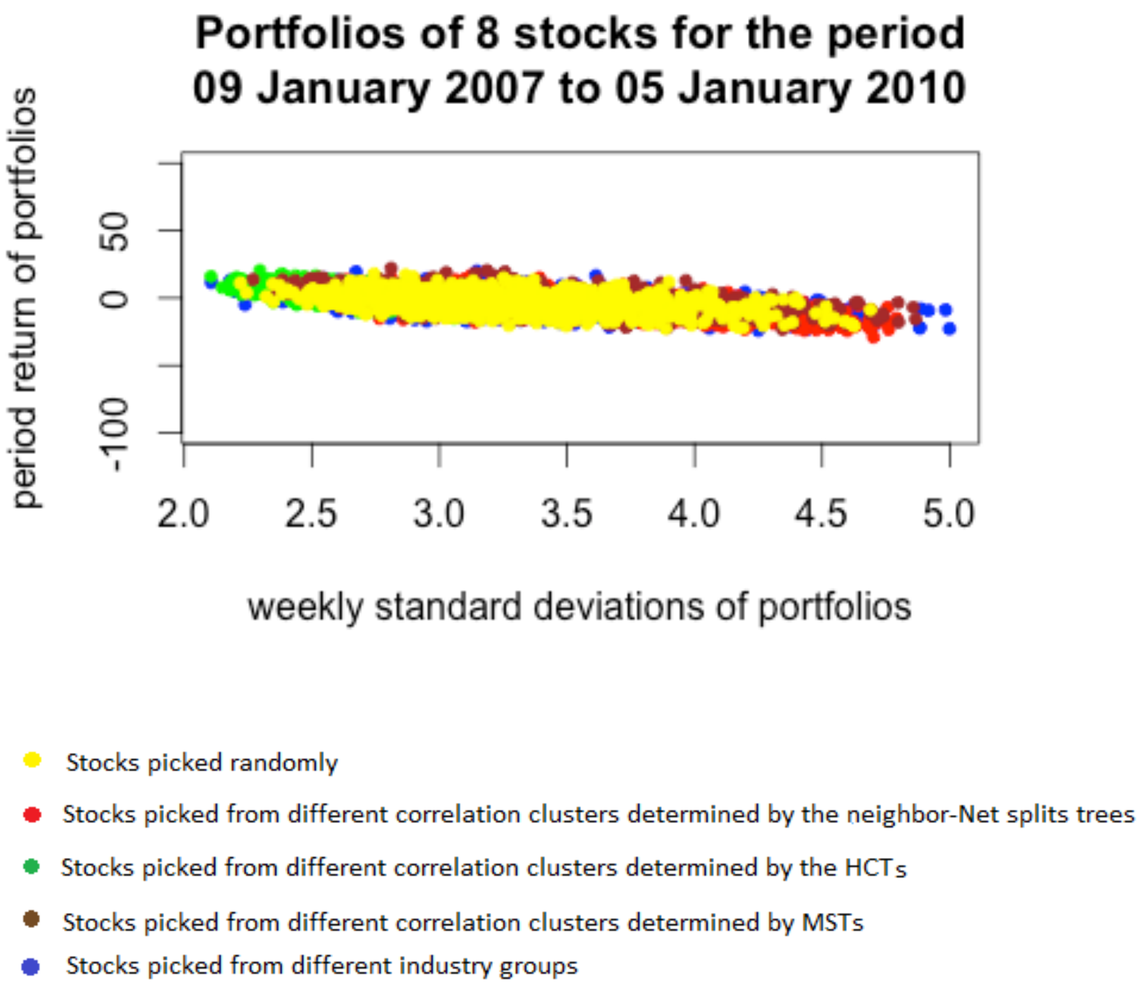


Figure 5.21 (f). Weekly standard deviation and return of each set of simulated 8-stock portfolios for period 5 using the five methods. The period returns and weekly standard deviations are expressed in percentage.

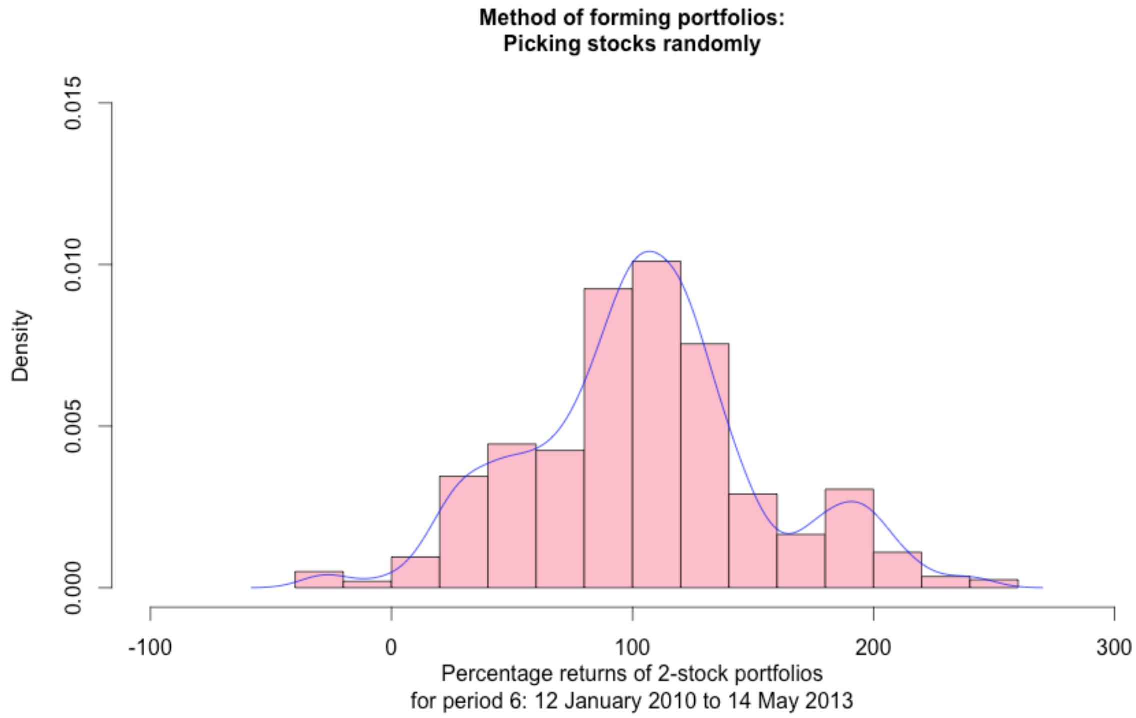


Figure 5.22 (a). The histogram and density curve of the simulated portfolios' percentage returns for period 6: 12 January 2010 to 14 May 2013. Each portfolio contains two stocks which were picked randomly.

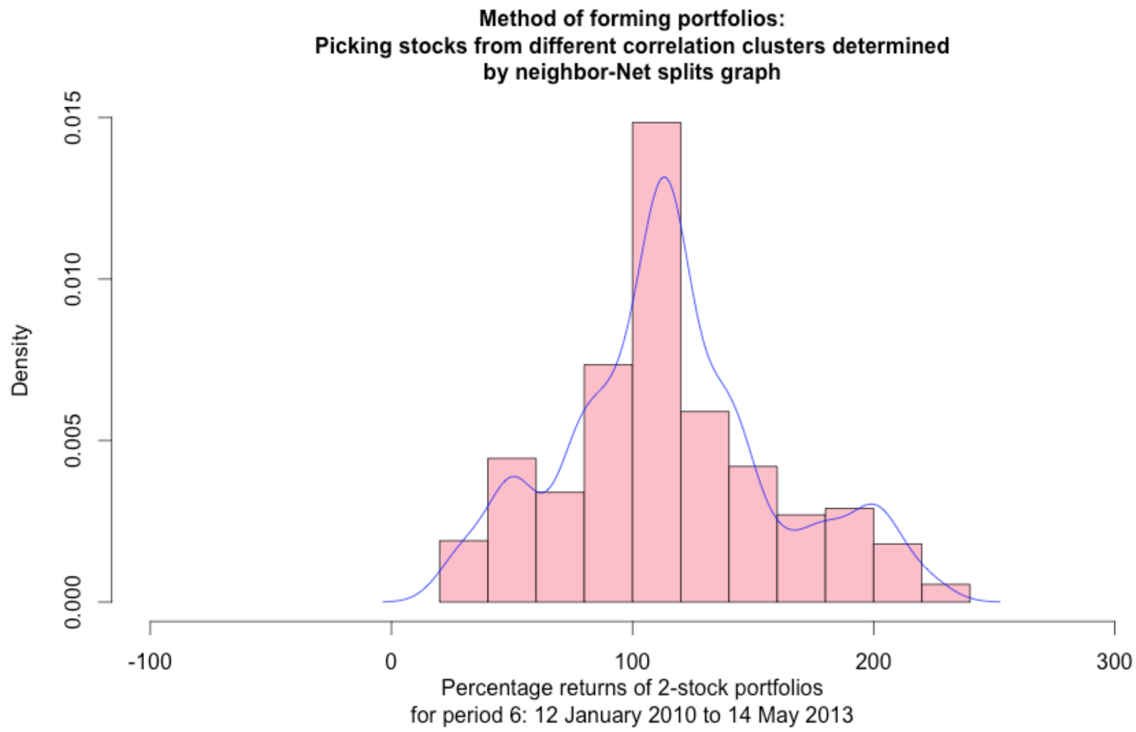


Figure 5.22 (b). The histogram and density curve of the simulated portfolios' percentage returns for period 6: 12 January 2010 to 14 May 2013. Each portfolio contains two stocks which were picked from different correlation clusters suggested by the neighbor-Net split graph produced from the stocks weekly returns in period 5: 09 January 2007 to 05 January 2010.

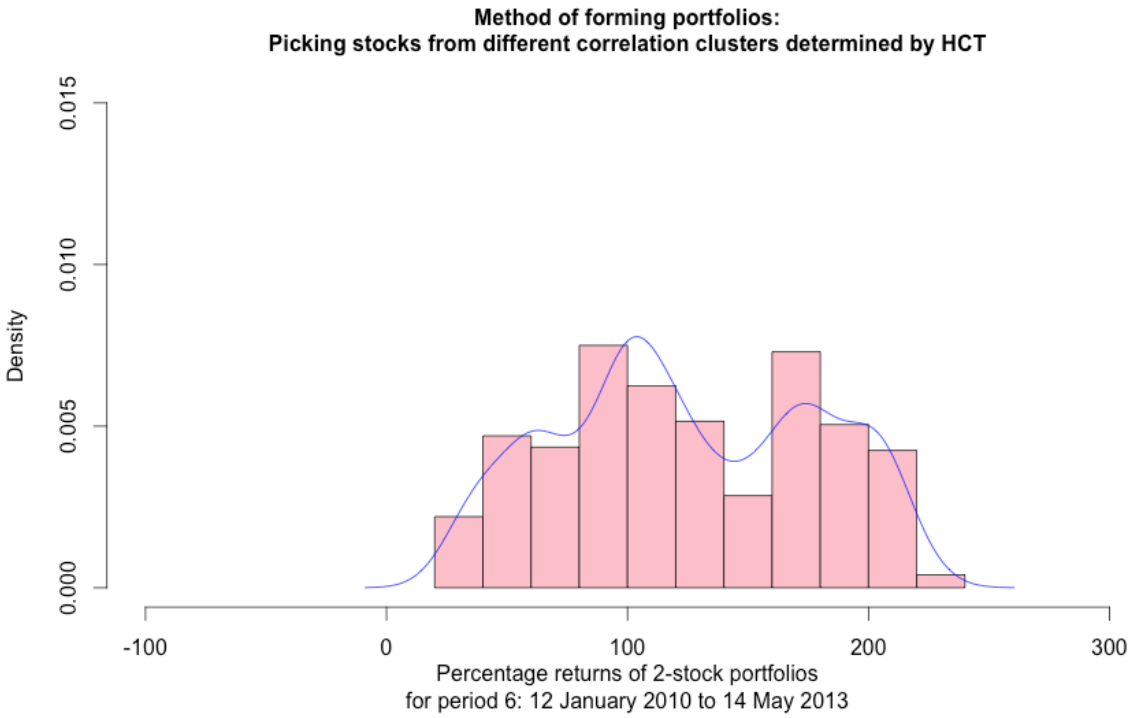


Figure 5.22 (c). The histogram and density curve of the simulated portfolios' percentage returns for period 6: 12 January 2010 to 14 May 2013. Each portfolio contains two stocks which were picked from different correlation clusters suggested by the HCT produced from the stocks weekly returns in period 5: 09 January 2007 to 05 January 2010.

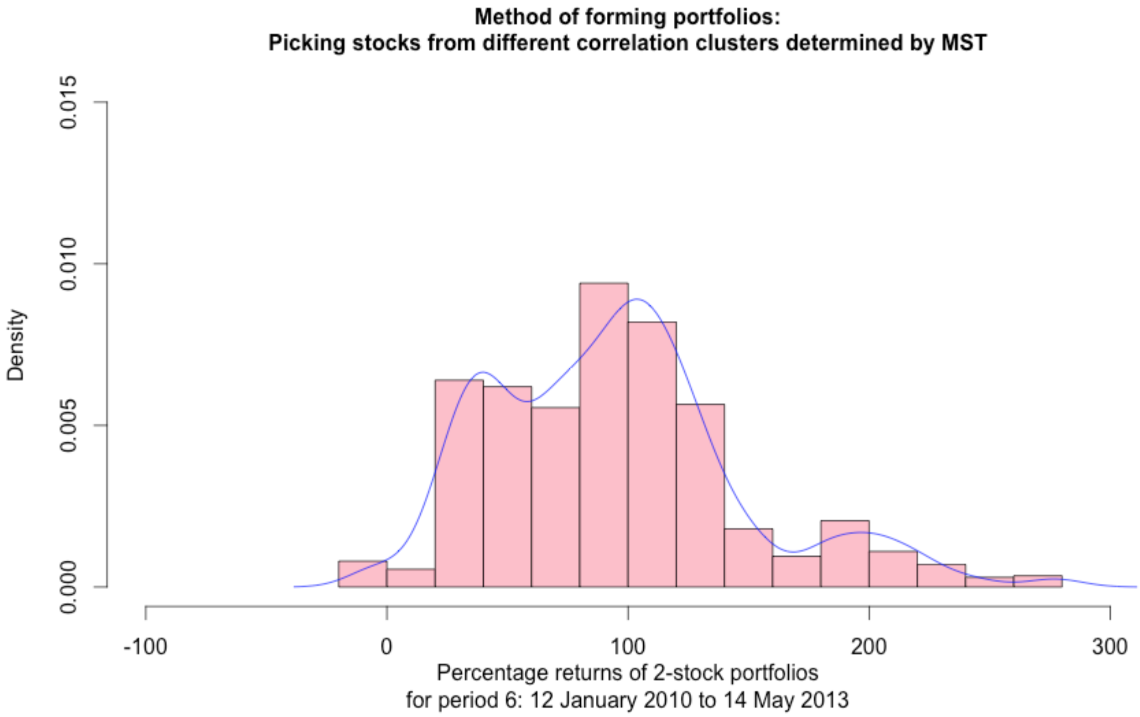


Figure 5.22 (d). The histogram and density curve of the simulated portfolios' percentage returns for period 6: 12 January 2010 to 14 May 2013. Each portfolio contains two stocks which were picked from different correlation clusters suggested by the MST produced from the stocks weekly returns in period 5: 09 January 2007 to 05 January 2010.

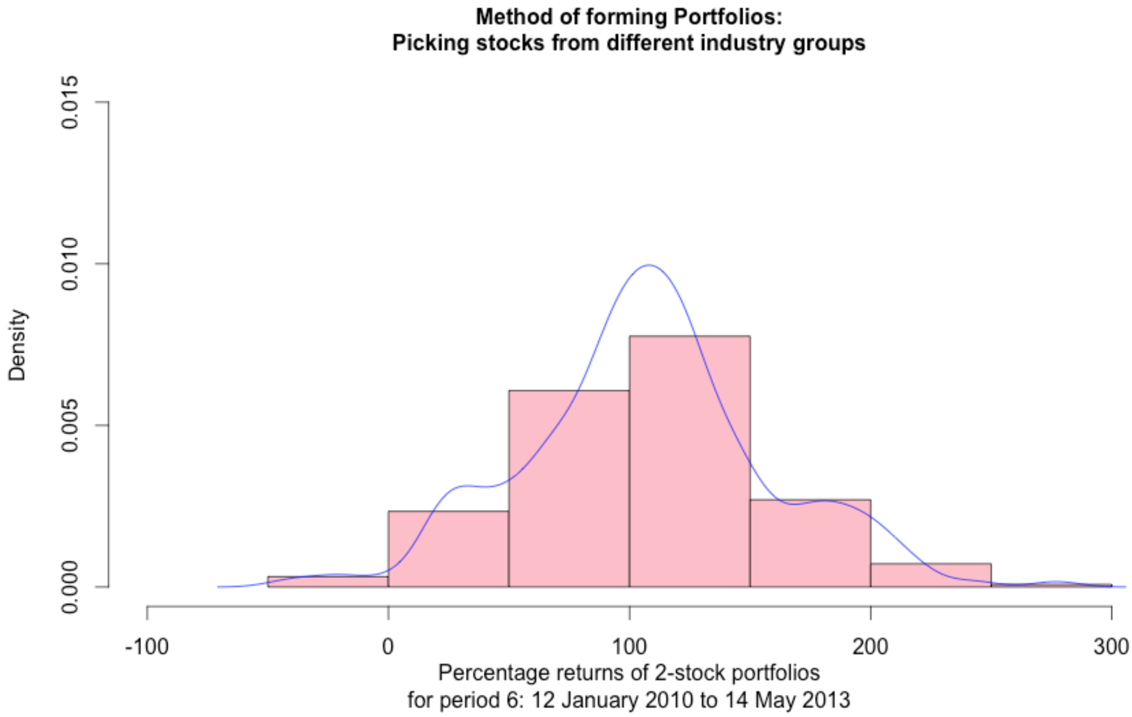


Figure 5.22 (e). The histogram and density curve of the simulated portfolios' percentage returns for period 6: 12 January 2010 to 14 May 2013. Each portfolio contains two stocks which were picked from different industry groups.

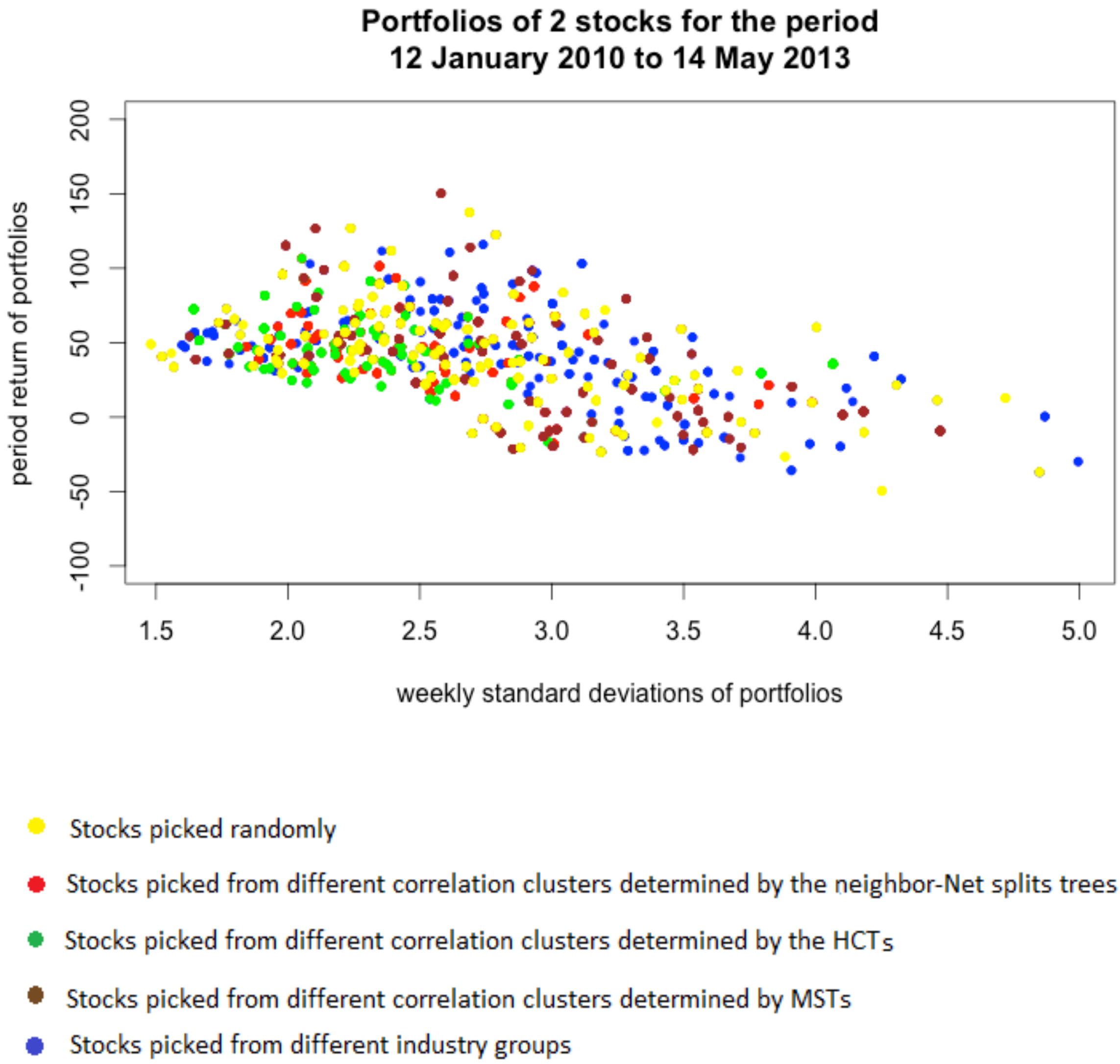


Figure 5.22 (f). Weekly standard deviation and return of each set of simulated 2-stock portfolios for period 6 using the five methods. The period returns and weekly standard deviations are expressed in percentage.

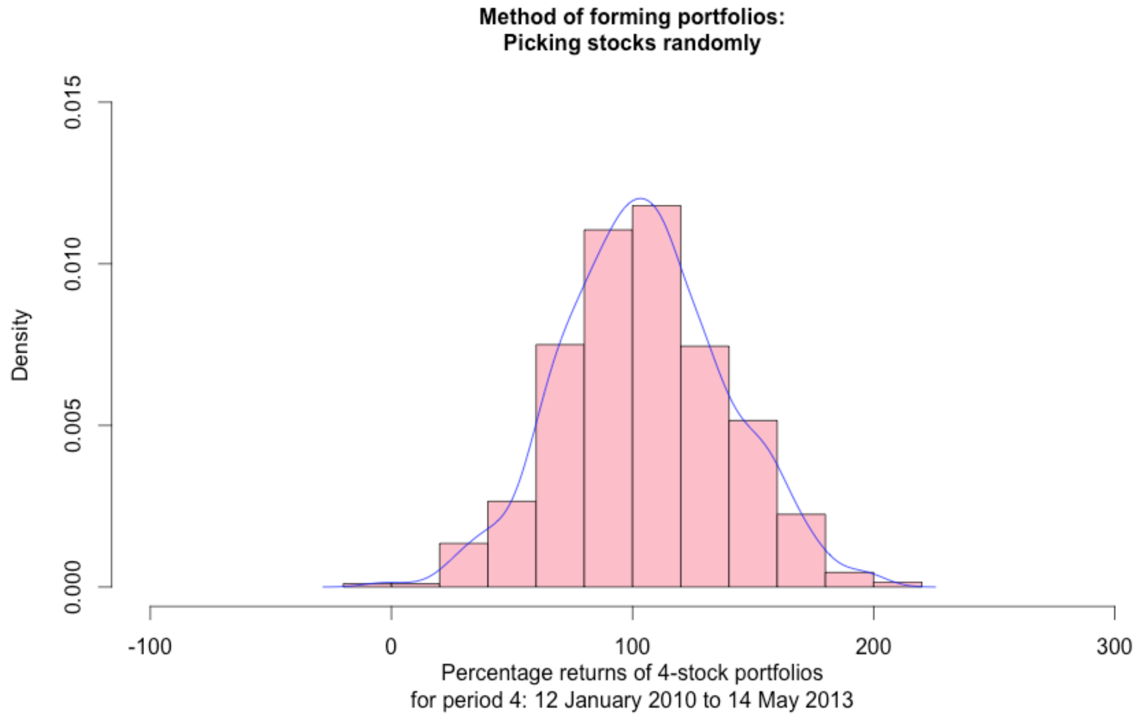


Figure 5.23 (a). The histogram and density curve of the simulated portfolios' percentage returns for period 6: 12 January 2010 to 14 May 2013. Each portfolio contains four stocks which were picked randomly.

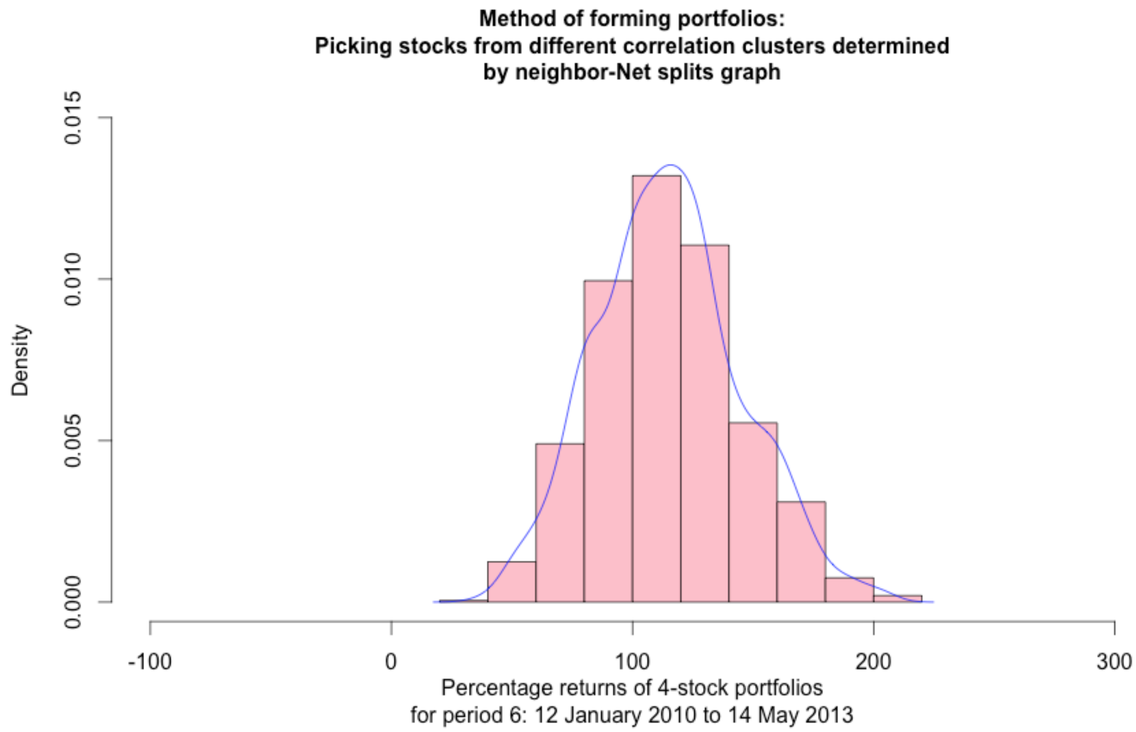


Figure 5.23 (b). The histogram and density curve of the simulated portfolios' percentage returns for period 6: 12 January 2010 to 14 May 2013. Each portfolio contains four stocks which were picked from different correlation clusters suggested by the neighbor-Net split graph produced from the stocks weekly returns in period 5: 09 January 2007 to 05 January 2010.

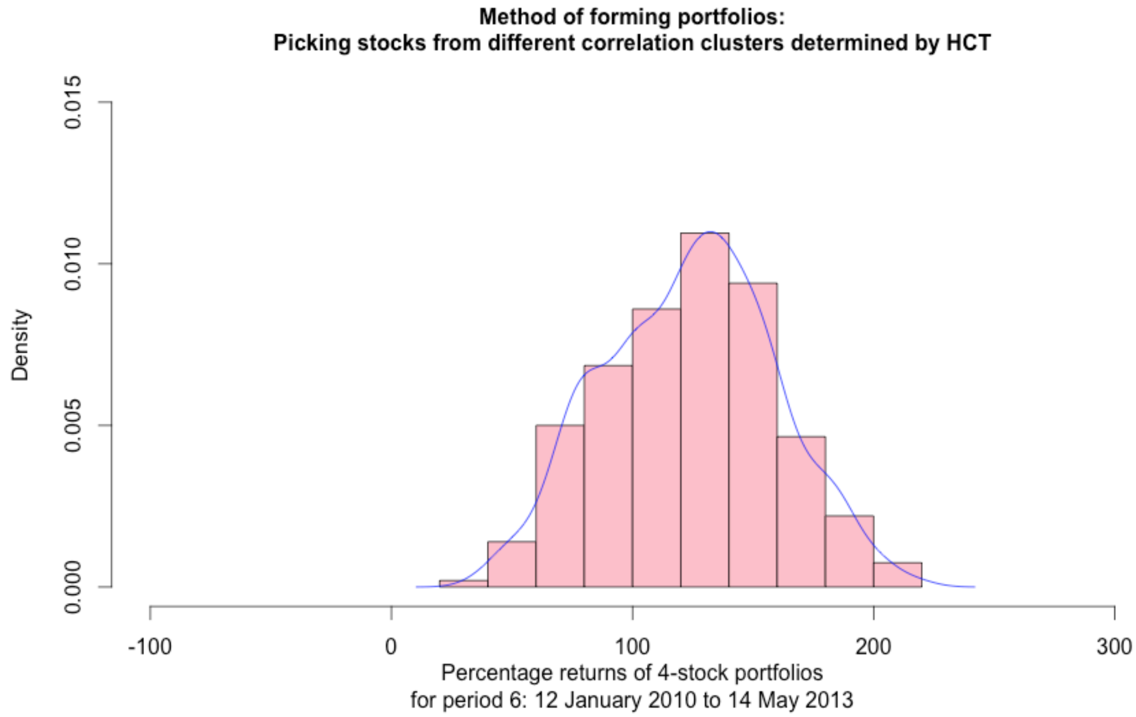


Figure 5.23 (c). The histogram and density curve of the simulated portfolios' percentage returns for period 6: 12 January 2010 to 14 May 2013. Each portfolio contains four stocks which were picked from different correlation clusters suggested by the HCT produced from the stocks weekly returns in period 5: 09 January 2007 to 05 January 2010.

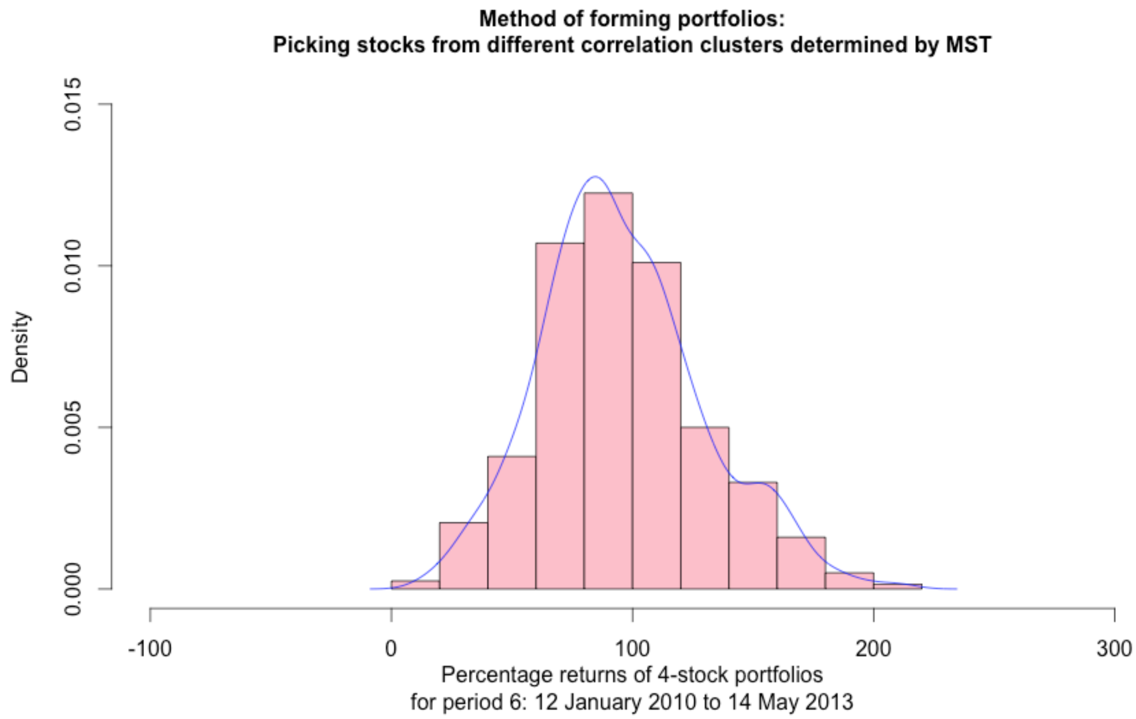


Figure 5.23 (d). The histogram and density curve of the simulated portfolios' percentage returns for period 6: 12 January 2010 to 14 May 2013. Each portfolio contains four stocks which were picked from different correlation clusters suggested by the MST produced from the stocks weekly returns in period 5: 09 January 2007 to 05 January 2010.

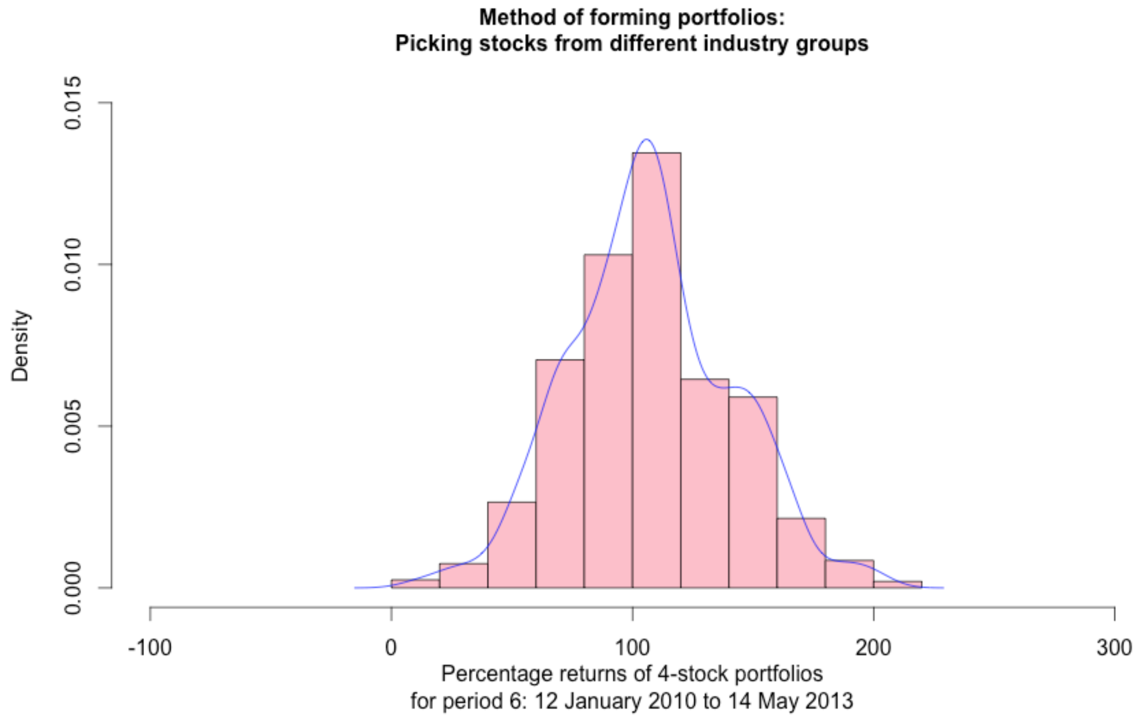


Figure 5.23 (e). The histogram and density curve of the simulated portfolios' percentage returns for period 6: 12 January 2010 to 14 May 2013. Each portfolio contains four stocks which were picked from different industry groups.

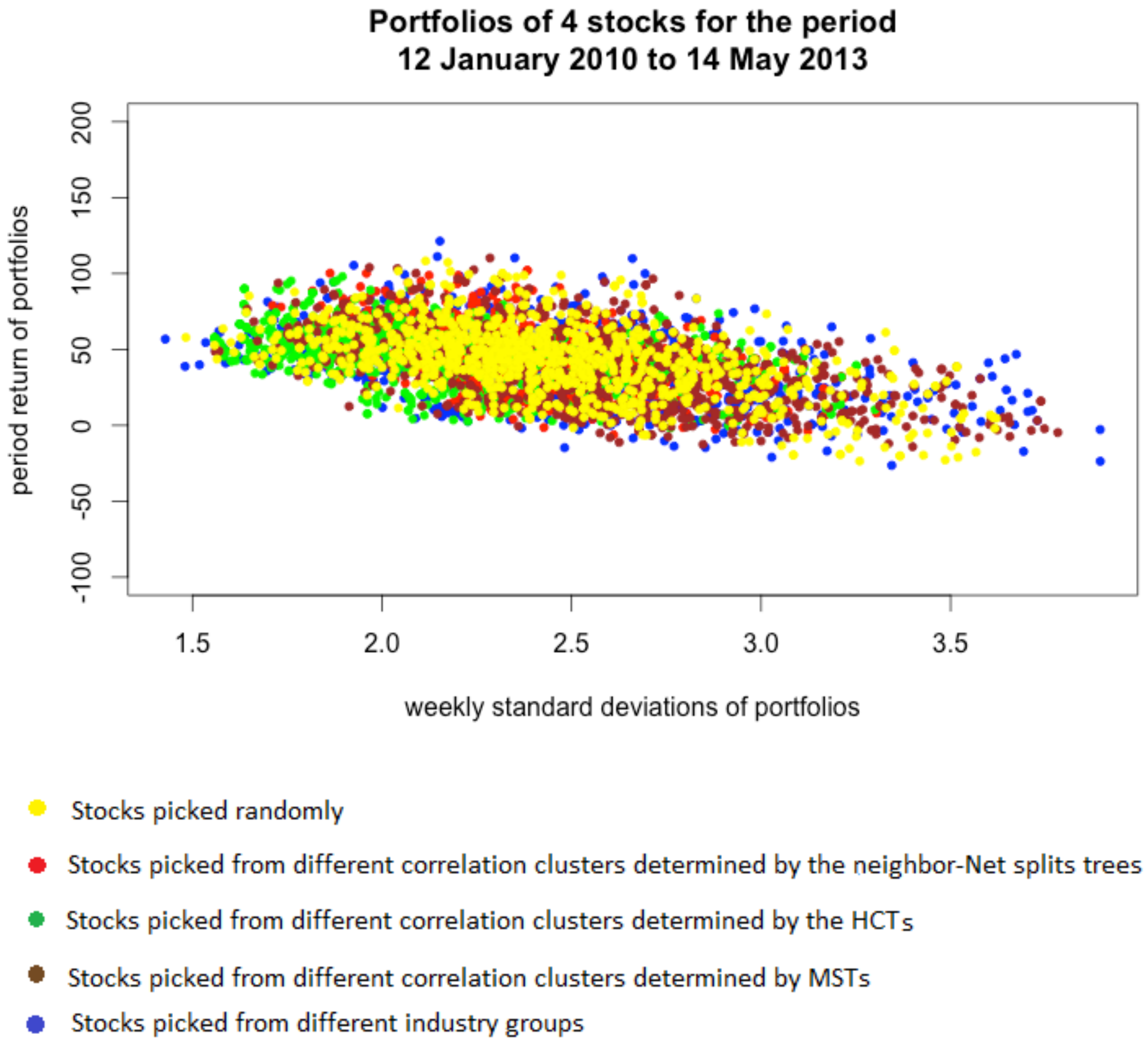


Figure 5.23 (f). Weekly standard deviation and return of each set of simulated 4-stock portfolios for period 6 using the five methods. The period returns and weekly standard deviations are expressed in percentage.

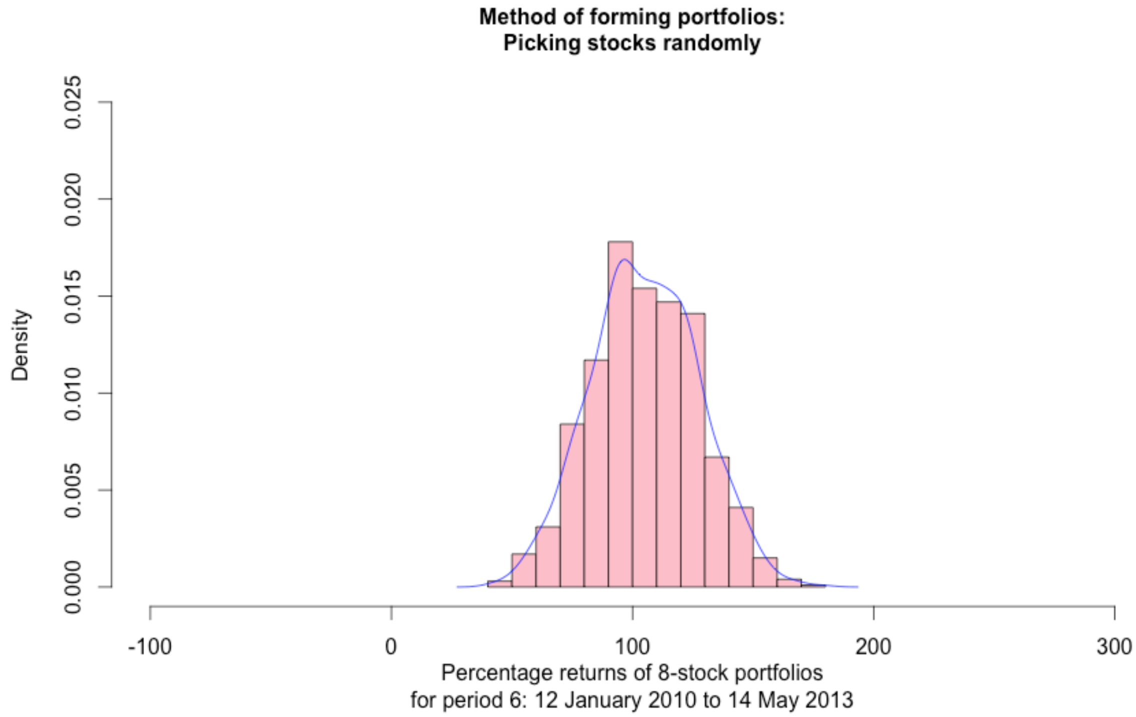


Figure 5.24 (a). The histogram and density curve of the simulated portfolios' percentage returns for period 6: 12 January 2010 to 14 May 2013. Each portfolio contains eight stocks which were picked randomly.

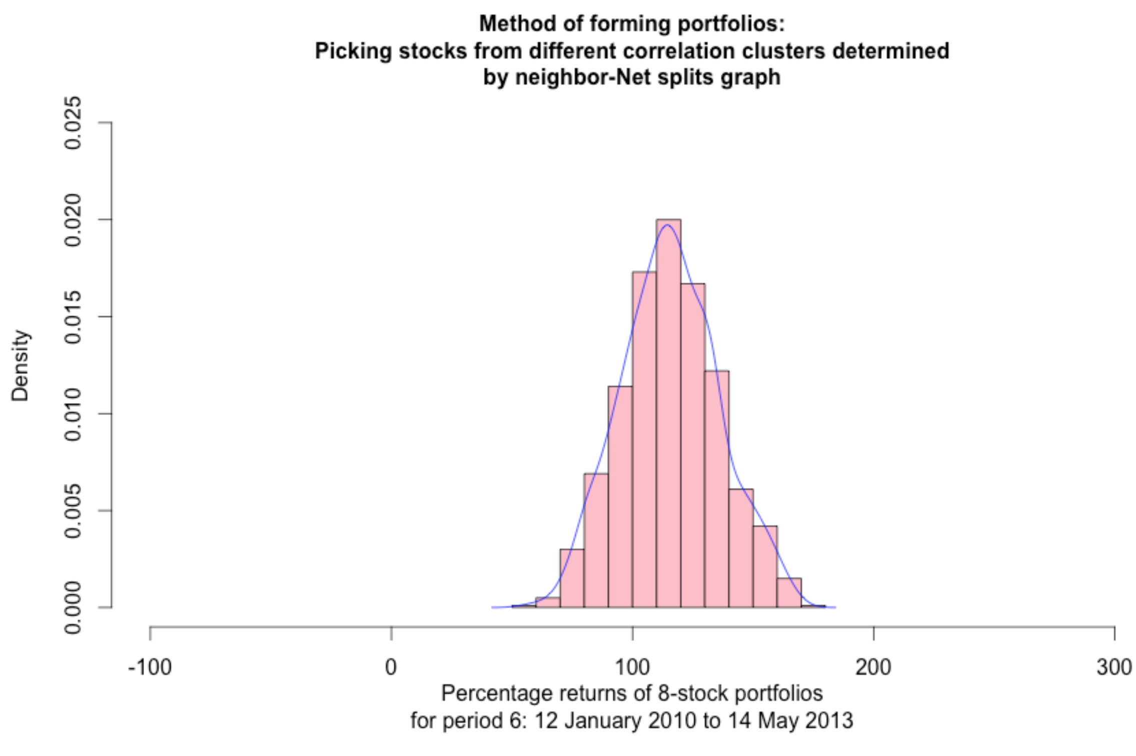


Figure 5.24 (b). The histogram and density curve of the simulated portfolios' percentage returns for period 6: 12 January 2010 to 14 May 2013. Each portfolio contains eight stocks which were picked from different correlation clusters suggested by the neighbor-Net split graph produced from the stocks weekly returns in period 5: 09 January 2007 to 05 January 2010.

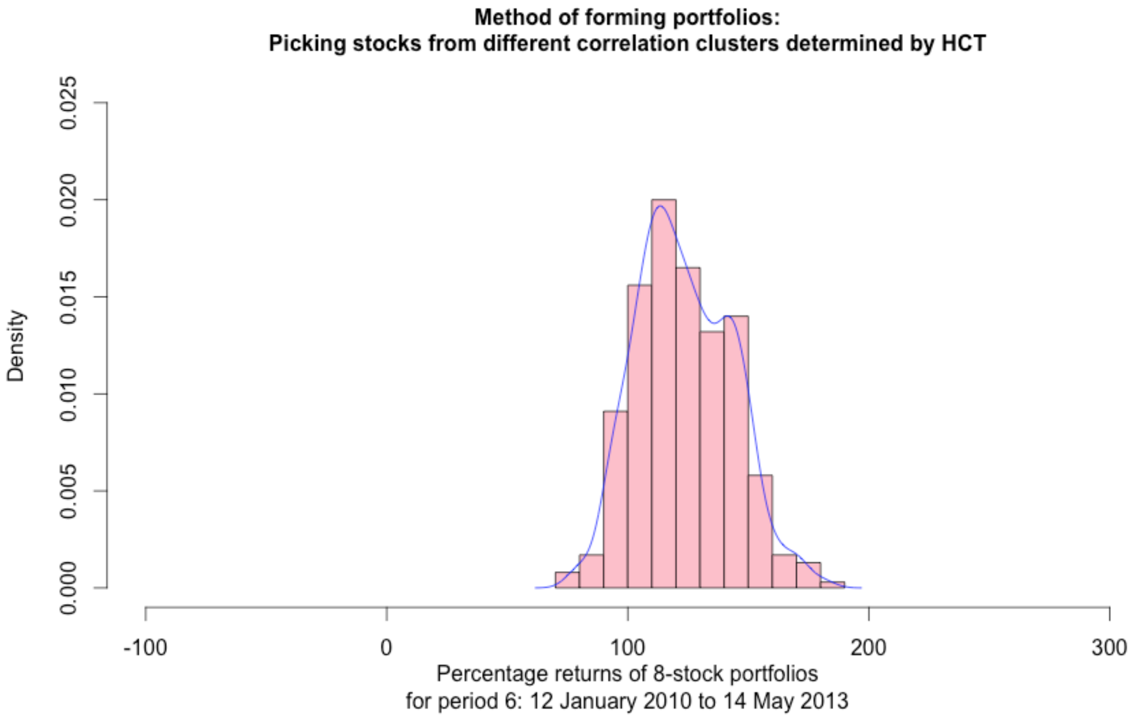


Figure 5.24 (c). The histogram and density curve of the simulated portfolios' percentage returns for period 6: 12 January 2010 to 14 May 2013. Each portfolio contains eight stocks which were picked from different correlation clusters suggested by the HCT produced from the stocks weekly returns in period 5: 09 January 2007 to 05 January 2010.

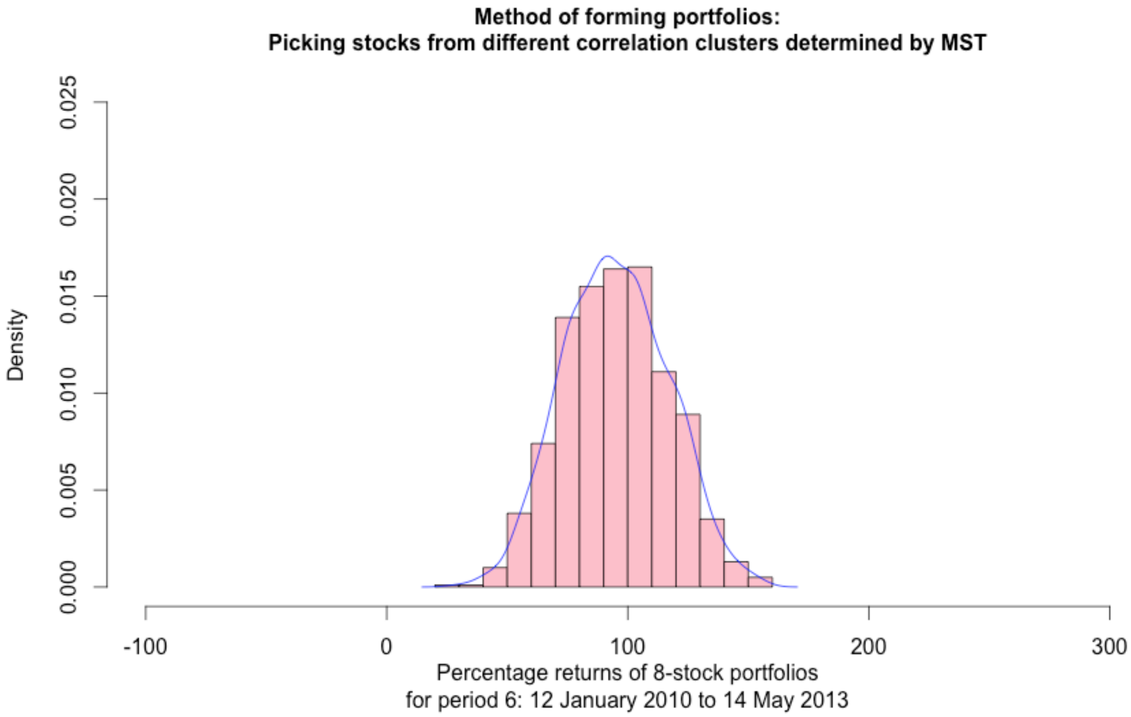


Figure 5.24 (d). The histogram and density curve of the simulated portfolios' percentage returns for period 6: 12 January 2010 to 14 May 2013. Each portfolio contains eight stocks which were picked from different correlation clusters suggested by the MST produced from the stocks weekly returns in period 5: 09 January 2007 to 05 January 2010.

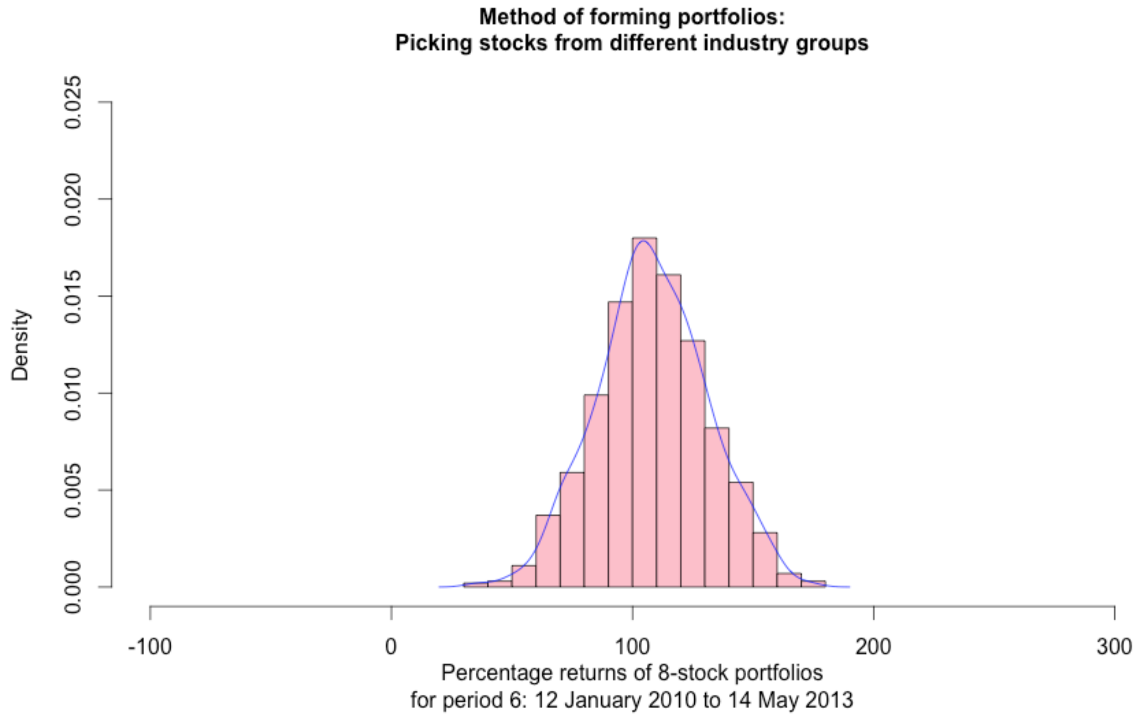


Figure 5.24 (e). The histogram and density curve of the simulated portfolios' percentage returns for period 6: 12 January 2010 to 14 May 2013. Each portfolio contains eight stocks which were picked from different industry groups.

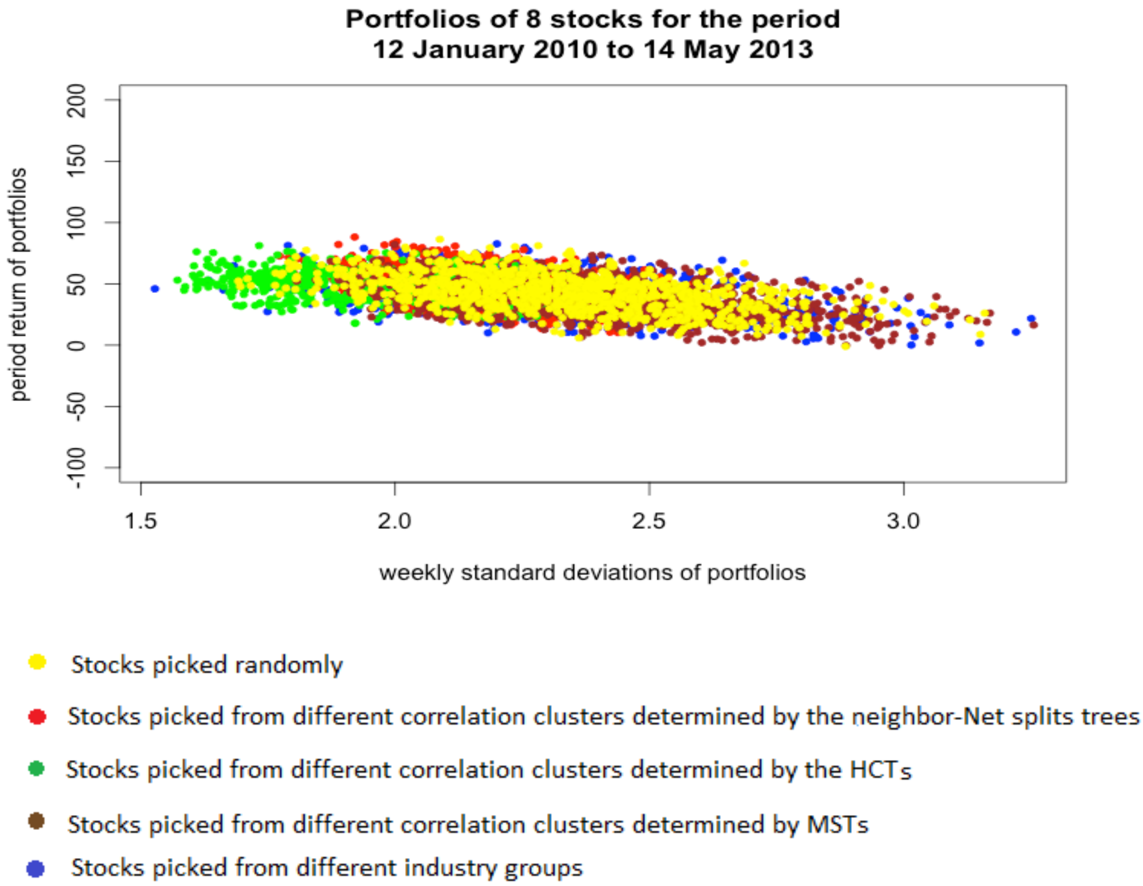


Figure 5.24 (f). Weekly standard deviation and return of each set of simulated 8-stock portfolios for period 6 using the five methods. The period returns and weekly standard deviations are expressed in percentage.

Appendix 7 Figures and Tables for Chapter 6

Period 2 clusters

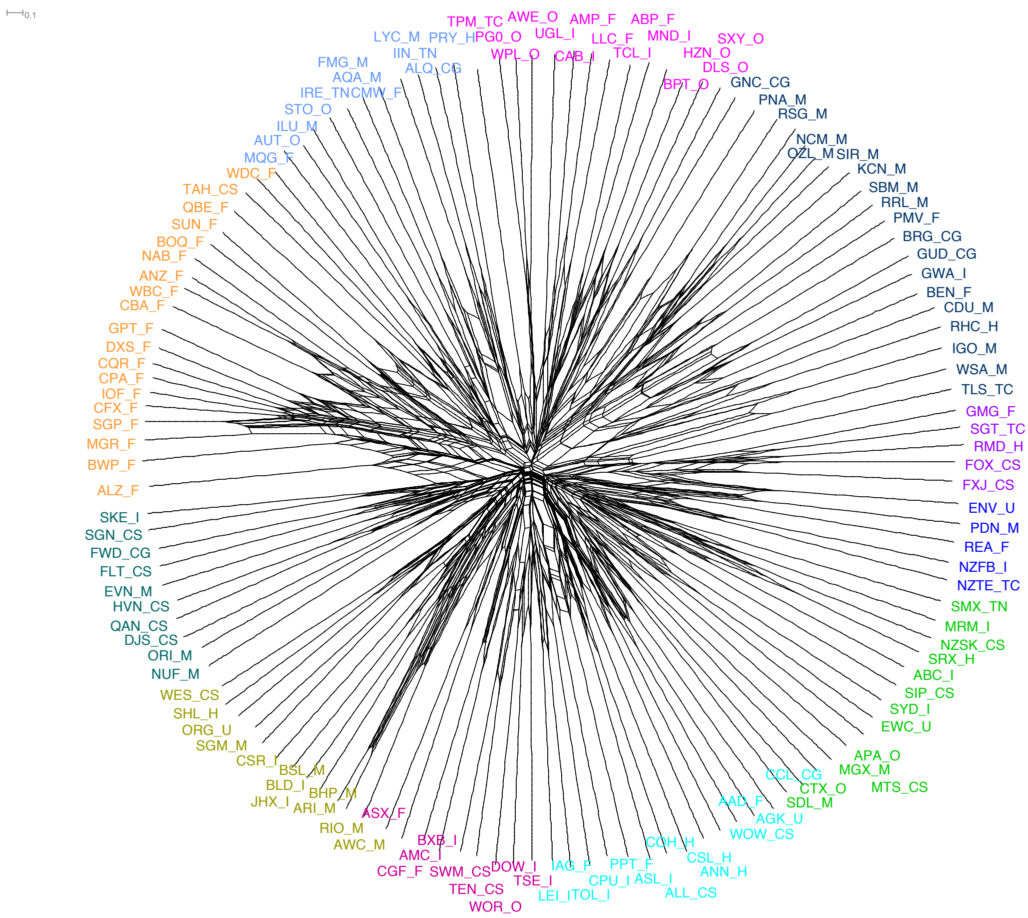


Figure 6.5. ASX 200 133 stocks in period 2: 2 April 2003 to 4 May 2005 is split into 11 correlation clusters suggested by neigh-Net splits graph.

Correlation Cluster Group	Stocks in each correlation cluster group
Cluster #1	GMG_F,SGT_TC,RMD_H,FOX_CS,FXJ_CS
Cluster #2	ENV_U,PDN_M,REA_F,NZFB_I,NZTE_TC
Cluster #3	SMX_TN,MRM_I,NZSK_CS,SRX_H,ABC_I,SIP_CS,SYD_I,MTS_CS,APA_O,MGX_M,CTX_O
Cluster #4	CCL_CG,AGK_U,WOW_CS,AAD_F,ANN_H,CSL_H,ALL_CS, COH_H,ASL_I,PPT_F,CPU_I,TOL_I,IAG_F,LEI_I
Cluster #5	TSE_I,DOW_I,WOR_O,TEN_CS,SWM_CS,BXB_I,AMC_I,CGF_F,ASX_F
Cluster #6	AWC_M,RIO_M,BHP_M,ARI_M,BSL_M,BLD_I,JHX_I,CSR_I, SGM_M,ORG_U,SHL_H,WES_CS
Cluster #7	NUF_M,ORI_M,DJS_CS,QAN_CS,HVN_CS,EVN_M,FLT_CS, FWD_CG,SGN_CS,SKE_I
Cluster #8	ALZ_F,BWP_F,MGR_F,SGP_F,CFX_F,IOF_F,CPA_F,CQR_F, DXS_F,GPT_F,CBA_F,WBC_F,ANZ_F,NAB_F,BOQ_F,SUN_F,QBE_F,TAH_CS,WDC_F
Cluster #9	MQG_F,AUT_O,ILU_M,STO_O,IRE_TN,AQA_M,CMW_F,LYC_M,IIN_TN,ALQ_CG,PRY_H
Cluster #10	TPM_TC, PG0_O, WPL_O, AWE_O, UGL_I, CAB_I, AMP_F, LLC_F, TCL_I, ABP_F, MND_I, BPT_O, HZN_O, DLS_O, SXY_O
Cluster #11	GNC_CG, PNA_M, RSG_M, OZL_M, NCM_M, SIR_M, KCN_M, SBM_M, RRL_M, PMV_F, BRG_CG, GUD_CG, GWA_I, BEN_F, RHC_H, IGO_M, WSA_M, TLS_TC

Table 6.10. The names of stocks in each correlation cluster for period 2: 2 April 2003 to 4 May 2005.

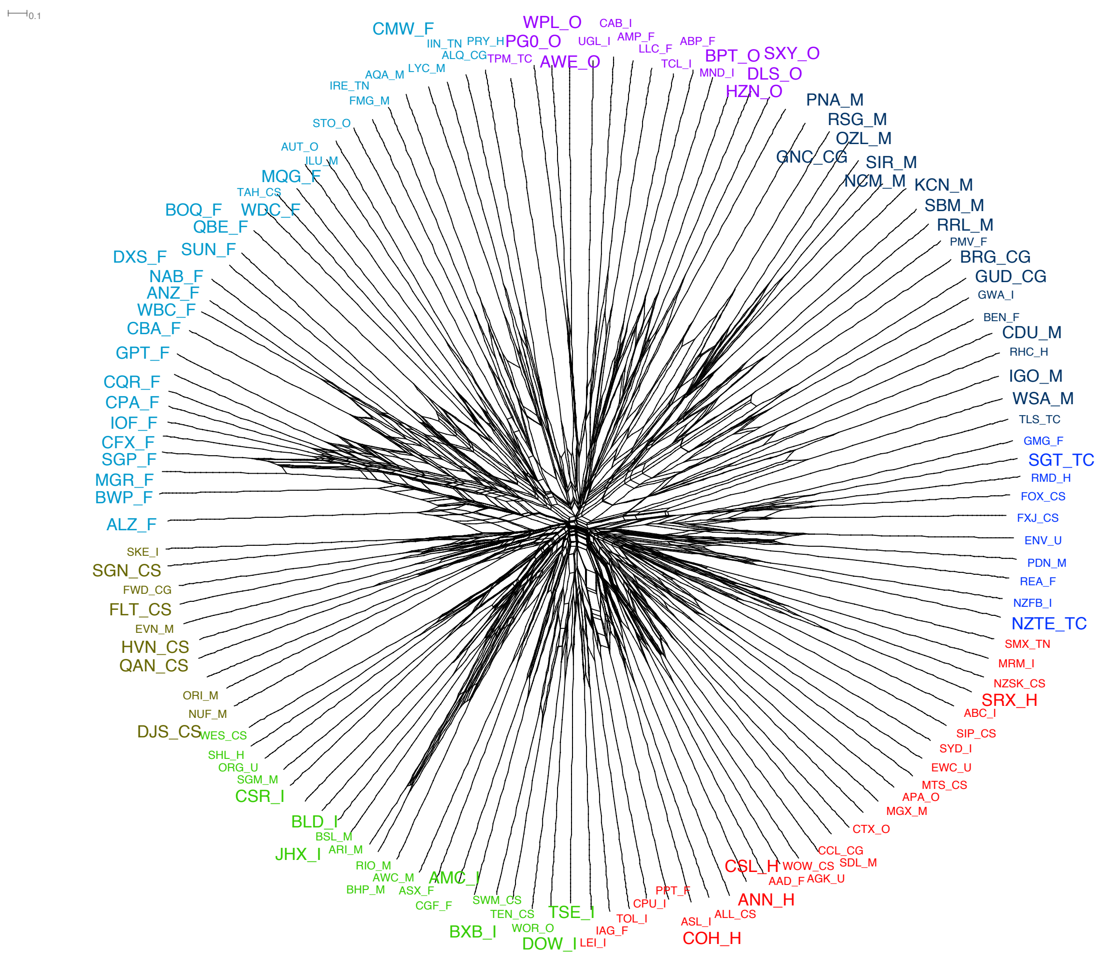


Figure 6.6. The ASX200's 133 stocks in period 2: 2 April 2003 to 4 May 2005 were split into total 14 clusters. Seven of the clusters were defined in stock picking strategy #4 (defined by colour and font size) and the other seven were defined in the stock picking strategy #5 (defined by colour and size).

Correlation with industry cluster group	Names of stocks in each cluster
Cluster #1	SGT_TC, NZTE_TC
Cluster #2	ANN_H, CSL_H, COH_H, SRX_H
Cluster #3	TSE_I, DOW_I, BXB_I, AMC_I, BLD_I, JHX_I, CSR_I
Cluster #4	DJS_CS, QAN_CS, HVN_CS, FLT_CS, SGN_CS
Cluster #5	ALZ_F, BWP_F, MGR_F, SGP_F, CFX_F, IOF_F, CPA_F, CQR_F, DXS_F, GPT_F, CBA_F, WBC_F, ANZ_F, NAB_F, BOQ_F, SUN_F, QBE_F, WDC_F, MQG_F, CMW_F
Cluster #6	PG0_O, WPL_O, AWE_O, BPT_O, HZN_O, DLS_O, SXY_O
Cluster #7	PNA_M, RSG_M, OZL_M, NCM_M, SIR_M, KCN_M, SBM_M, RRL_M, IGO_M, WSA_M, GNC_CG, BRG_CG, GUD_CG

Table 6.11. The code of each stock within each industry plus correlation group for period 2: 2 April 2003 to 4 May 2005.

Correlation without industry cluster group	Names of stocks in each cluster
Cluster #1	GMG_F, RMD_H, FOX_CS, FXJ_CS, ENV_U, PDN_M, REA_F, NZFB_I
Cluster #2	SMX_TN, MRM_I, NZSK_CS, ABC_I, SIP_CS, SYD_I, MTS_CS, APA_O, MGX_M, CTX_O, CCL_CG, AGK_U, WOW_CS, AAD_F, ALL_CS, ASL_I, PPT_F, CPU_I, TOL_I, IAG_F, LEI_I
Cluster #3	WOR_O, TEN_CS, SWM_CS, CGF_F, ASX_F, AWC_M, RIO_M, BHP_M, ARI_M, BSL_M, SGM_M, ORG_U, SHL_H, WES_CS
Cluster #4	NUF_M, ORI_M, EVN_M, FWD_CG, SKE_I
Cluster #5	TAH_CS, AUT_O, ILU_M, STO_O, IRE_TN, AQA_M, LYC_M, IIN_TN, ALQ_CG, PRY_H
Cluster #6	TPM_TC, UGL_I, CAB_I, AMP_F, LLC_F, TCL_I, ABP_F, MND_I
Cluster #7	PMV_F, GWA_I, BEN_F, RHC_H, TLS_TC

Table 6.12. The code of each stock within each non-industry plus correlation group for period 2: 2 April 2003 to 4 May 2005.

Period 3 Cluster

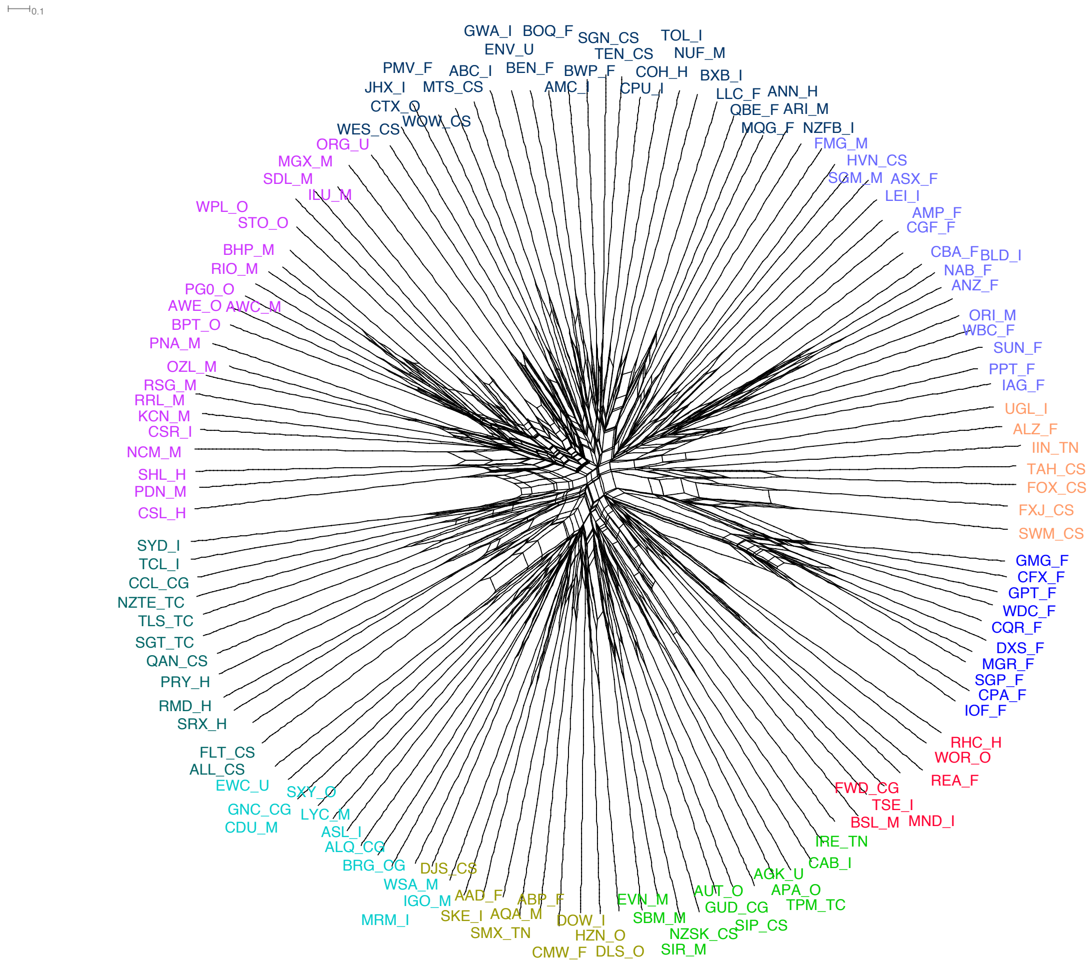


Figure 6.7. ASX 200 133 stocks in period 3:11 May 2005 to 10 October 2007 is split into 10 correlation clusters suggested by neigh-Net splits graph.

Correlation Cluster Group	Stocks in each correlation cluster group
Cluster #1	GMG_F, CFX_F, GPT_F, WDC_F, CQR_F, DXS_F, MGR_F, SGP_F, CPA_F, IOF_F
Cluster #2	RHC_H, WOR_O, REA_F, MND_I, TSE_I, BSL_M, FWD_CG
Cluster #3	IRE_TN, CAB_I, TPM_TC, APA_O, AGK_U, SIP_CS, GUD_CG, AUT_O, NZSK_CS, SIR_M, SBM_M, EVN_M
Cluster #4	DLS_O, HZN_O, DOW_I, CMW_F, ABP_F, AQA_M, SMX_TN, AAD_F, SKE_I, DJS_CS
Cluster #5	IGO_M, WSA_M, MRM_I, BRG_CG, ALQ_CG, ASL_I, LYC_M, SXY_O, GNC_CG, CDU_M, EWC_U
Cluster #6	FLT_CS, ALL_CS, SRX_H, RMD_H, PRY_H, QAN_CS, SGT_TC, TLS_TC, NZTE_TC, CCL_CG, TCL_I, SYD_I
Cluster #7	CSL_H, PDN_M, SHL_H, NCM_M, CSR_I, KCN_M, RRL_M, RSG_M, OZL_M, PNA_M, BPT_O, AWE_O, PG0_O, WPL_O, RIO_M, BHP_M, AWC_M, STO_O, SDL_M, MGX_M, ILU_M, ORG_U
Cluster #8	WES_CS, JHX_I, CTX_O, PMV_F, WOW_CS, MTS_CS, ABC_I, GWA_I, ENV_U, BEN_F, BOQ_F, AMC_I, BWP_F, SGN_CS, TEN_CS, CPU_I, COH_H, TOL_I, NUF_M, BXB_I, LLC_F, QBE_F, MQG_F, ANN_H, ARI_M, NZFB_I
Cluster #9	FMG_M, SGM_M, HVN_CS, LEI_I, ASX_F, CGF_F, AMP_F, CBA_F, NAB_F, ANZ_F, WBC_F, ORI_M, BLD_I, SUN_F, PPT_F, IAG_F
Cluster #10	UGL_I, ALZ_F, IIN_TN, TAH_CS, FOX_CS, FXJ_CS, SWM_CS

Table 6.13. The names of stocks in each correlation cluster for period 3: 11 May 2005 to 10 October 2007.

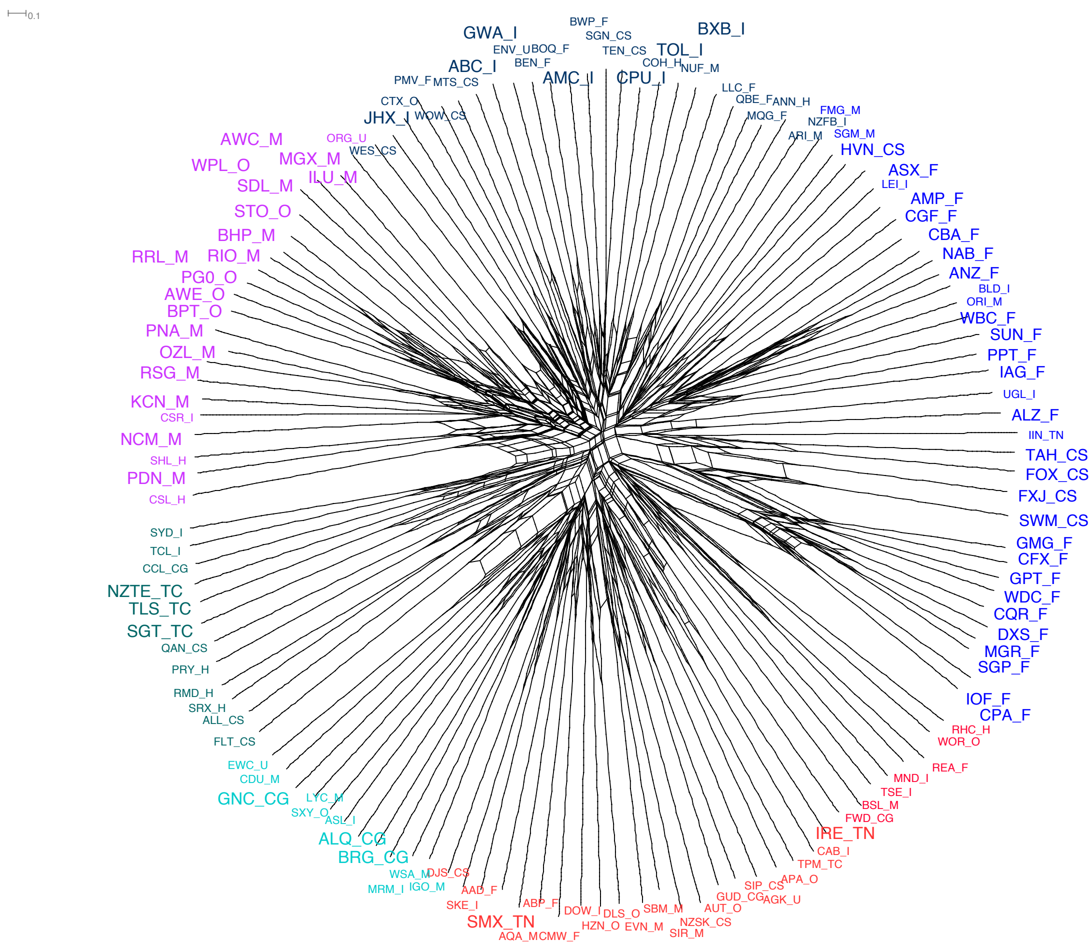


Figure 6.8. The ASX200's 133 stocks in period 3: 11 May 2005 to 10 October 2007 were split into total 12 clusters. Six of the clusters were defined in stock picking strategy #4 (defined by colour and font size) and the other six were defined in the stock picking strategy #5 (defined by colour and size).

Correlation with industry cluster group	Stocks in each cluster
Cluster #1	ASX_F,CGF_F,AMP_F,CBA_F,NAB_F,ANZ_F,WBC_F,SUN_F,PPT_F,IAG_F,GMG_F,CFX_F,GPT_F,WDC_F,CQR_F,DXS_F,MGR_F,SGP_F,CPA_F,ALZ_F,IOF_F,HVN_CS,TAH_CS,FOX_CS,FXJ_CS,SWM_CS
Cluster #2	IRE_TN,SMX_TN
Cluster #3	BRG_CG,ALQ_CG,GNC_CG
Cluster #4	SGT_TC,TLS_TC,NZTE_TC
Cluster #5	PDN_M,NCM_M,KCN_M,RRL_M,RSG_M,OZL_M,PNA_M,RIO_M,BHP_M,AWC_M,SDL_M,MGX_M,ILU_M,BPT_O,AWE_O,PG0_O,WPL_O,STO_O
Cluster #6	JHX_I,ABC_I,GWA_I,AMC_I,CPU_I,TOL_I,BXB_I,NZFB_I

Table 6.14. The code of each stock within each industry plus correlation group for period 3.

Correlation without industry cluster group	Stocks in each cluster
Cluster #1	FMG_M,SGM_M,LEI_I,ORI_M,BLD_I,UGL_I,IIN_TN
Cluster #2	RHC_H,WOR_O,REA_F,MND_I,TSE_I,BSL_M,FWD_CG,CAB_I,TPM_TC,APA_O,AGK_U,SIP_CS,GUD_CG,AUT_O,NZSK_CS,SIR_M,SBM_M,EVN_M,DLS_O,HZN_O,DOW_I,CMW_F,ABP_F,AQA_M,AAD_F,SKE_I,DJS_CS
Cluster #3	IGO_M,WSA_M,MRM_I,ASL_I,LYC_M,SXY_O,CDU_M,EWC_U
Cluster #4	FLT_CS,ALL_CS,SRX_H,RMD_H,PRY_H,QAN_CS,CCL_CG,TCL_I,SYD_I
Cluster #5	CSL_H,SHL_H,CSR_I,ORG_U
Cluster #6	WES_CS,CTX_O,PMV_F,WOW_CS,MTS_CS,ENV_U,BEN_F,BOQ_F,BWP_F,SGN_CS,TEN_CS,COH_H,NUF_M,LLC_F,QBE_F,MQG_F,ANN_H,ARI_M

Table 6.15. The code of each stock within each non-industry plus correlation group for period 3.

Period 4 clusters

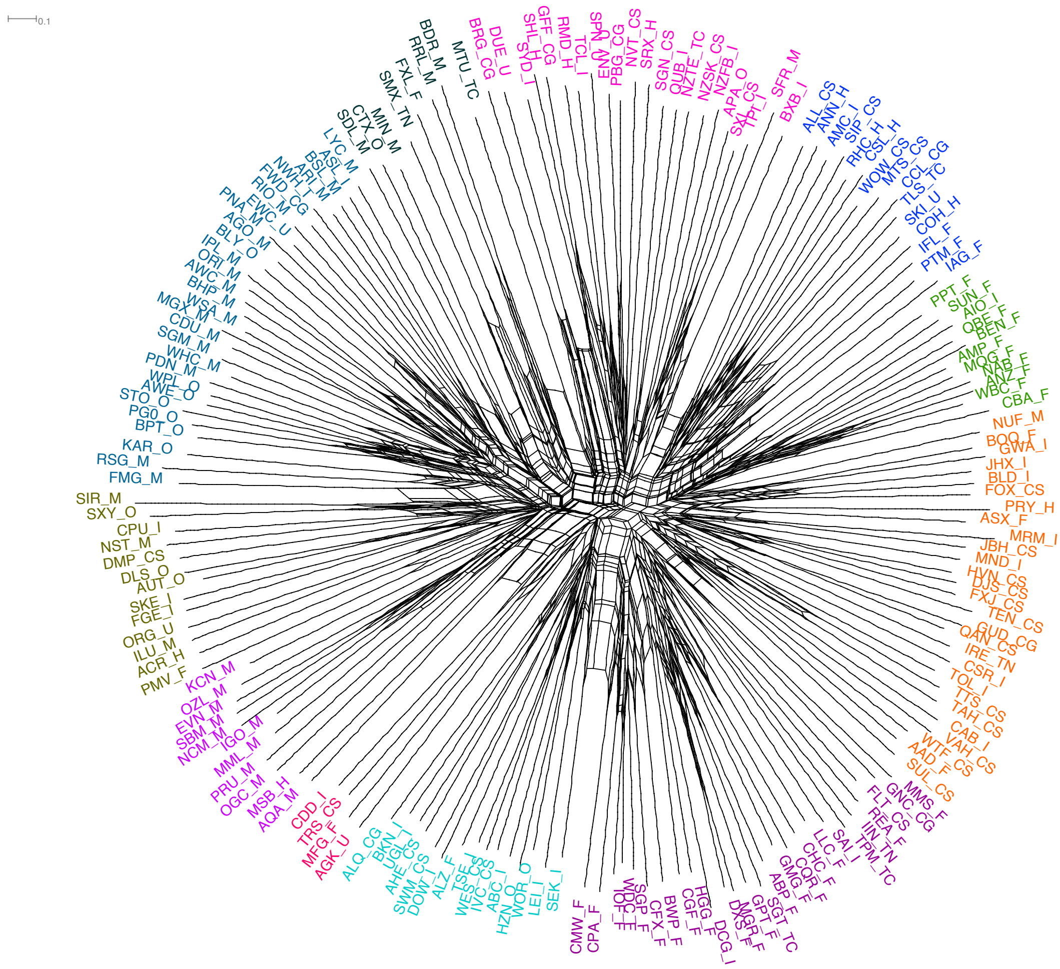


Figure 6.9. ASX 200 181 stocks in period 4: 17 October 2007 to 04 March 2009 is split into 11 correlation clusters suggested by neigh-Net splits graph.

Correlation Cluster Group	Stocks in each correlation cluster group
Cluster #1	ALL_CS,ANN_H,AMC_I,SIP_CS,RHC_H,CSL_H,WOW_CS, MTS_CS,CCL_CG,TLS_TC,SKI_U,COH_H,IFL_F,PTM_F,IAG_F
Cluster #2	PPT_F,SUN_F,AIO_I,QBE_F,BEN_F,AMP_F,MQG_F,NAB_F,ANZ_F,WBC_F,CBA_F
Cluster #3	NUF_M,BOQ_F,GWA_I,JHX_I,BLD_I,FOX_CS,PRY_H,ASX_F,MRM_I,JBH_CS,MND_I,HVN_CS,DJS_CS,FXJ_CS,TEN_CS,GUD_CG,QAN_CS,IRE_TN,CSR_I,TOL_I,TTS_CS,TAH_CS,CAB_I,VAH_CS, WTF_CS,AAD_F,SUL_CS
Cluster #4	MMS_F,GNC_CG,FLT_CS,REA_F,IIN_TN,SAI_I,LLC_F,CHC_F,CQR_F,GMG_F,ABP_F,SGT_TC,GPT_F,MGR_F, DXS_F,DCG_I,HGG_F,CGF_F,BWP_F CFX_F,SGP_F,WDC_F,IOF_F,CPA_F,CMW_F
Cluster #5	SEK_I,LEI_I,WOR_O,HZN_O,ABC_I,IVC_CS,WES_CS,TSE_I,ALZ_F,DOW_I,SWM_CS,AHE_CS,UGL_I,BKN_I,ALQ_CG
Cluster #6	AGK_U,MFG_F,TRS_CS,CDD_I
Cluster #7	AQA_M,MSB_H,OGC_M, PRU_M,MML_M,IGO_M,NCM_M,SBM_M,EVN_M,OZL_M, KCN_M
Cluster #8	PMV_F,ACR_H,ILU_M,ORG_U,SKE_I,DLS_O,DMP_CS,CPU_I,SXY_O,SIR_M
Cluster #9	FMG_M,RSG_M,KAR_O,BPT_O,PG0_O,STO_O,AWE_O,WPL_O,PDN_M,WHC_M,SGM_M,CDU_M,MGX_M,WSA_M,BHP_M,AWC_M,ORI_M,IPL_M,BLY_O,AGO_M,PNA_M,EWC_U,RIO_M,FWD_CG,ARI_M,BSL_M,ASL_I,LYC_M
Cluster #10	SDL_M,CTX_O,MIN_M,SMX_TN,FXL_F,RRL_M,BDR_M,MTU_TC
Cluster #11	BRG_CG,DUE_U,SYD_I,SHL_H,GFF_CG,RMD_H,TCL_I,SPN_U,ENV_U,PBG_CG,NVT_CS,SRX_H,SGN_CS,QUB_I,NZTE_TC,NZSK_CS,NZFB_I,APA_O,SXL_CS,TPI_I,BXB_I

Table 6.16. The names of stocks in each correlation cluster for period 4: 17 October 2007 to 04 March 2009.

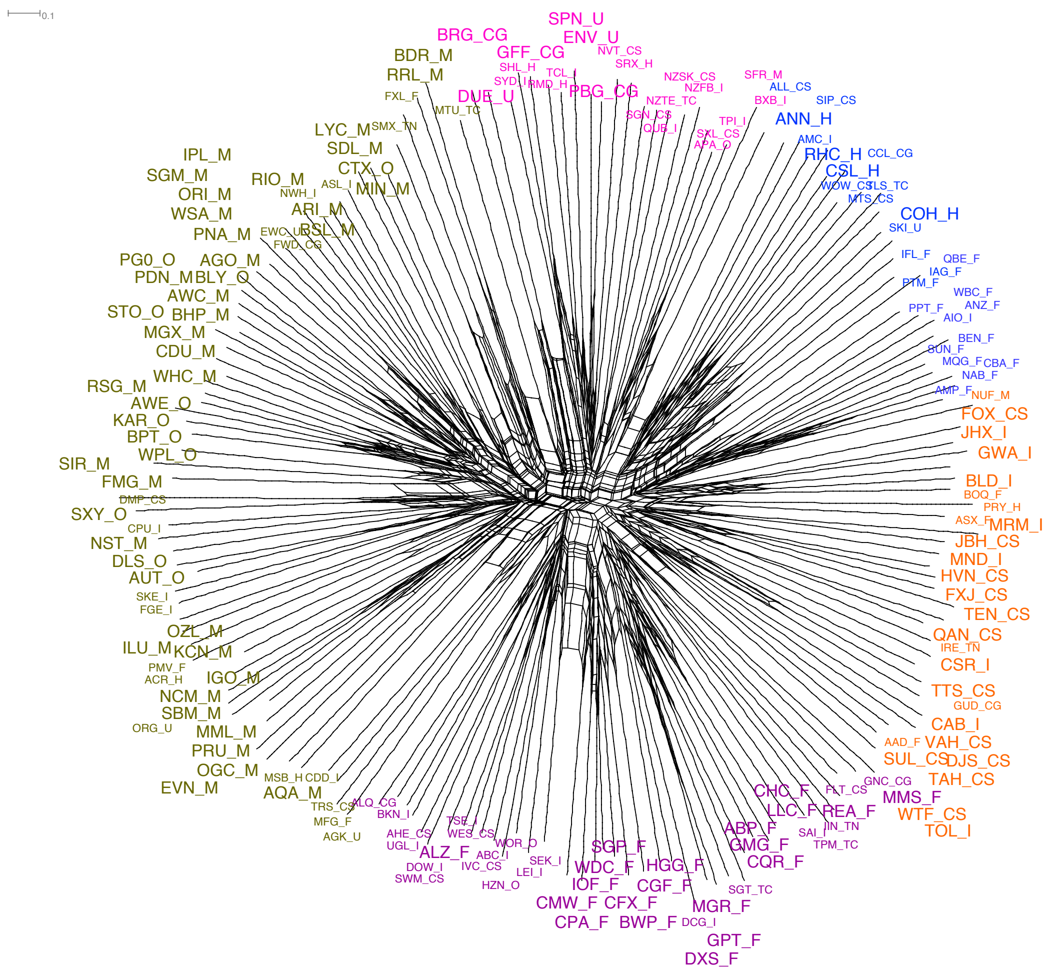


Figure 6.10. The ASX200's 181 stocks in period 4: 17 October 2007 to 04 March 2009 were split into total 12 clusters. Six of the clusters were defined in stock picking strategy #4 (defined by colour and font size) and the other six were defined in the stock picking strategy #5 (defined by colour and size).

Correlation with industry cluster group	Stocks in each cluster
Cluster #1	ANN_H,RHC_H,CSL_H,COH_H
Cluster #2	GWA_I,JHX_I,BLD_I,FOX_CS,MRM_I,JBH_CS,MND_I,HVN_CS,DJS_CS,FXJ_CS,TEN_CS,QAN_CS,CSR_I,TOL_I,TTS_CS,TAH_CS,CAB_I,VAH_CS,WTF_CS,SUL_CS
Cluster #3	MMS_F,REA_F,LLC_F,CHC_F,CQR_F,GMG_F,ABP_F,GPT_F,MGR_F,DXS_F,HGG_F,CGF_F,BWP_F,CFX_F,SGP_F,WDC_F,IOF_F,CPA_F,CMW_F,ALZ_F
Cluster #4	AQA_M,OGC_M,PRU_M,MML_M,IGO_M,NCM_M,SBM_M,EVN_M,OZL_M,KCN_M,ILU_M,DLS_O,SXY_O,SIR_M,FMG_M,RSG_M,KAR_O,BPT_O,PG0_O,STO_O,AWE_O,WPL_O,PDN_M,WHC_M,SGM_M,CDU_M,MGX_M,WSA_M,BHP_M,AWC_M,ORI_M,IPL_M,BLY_O,AGO_M,PNA_M,RIO_M,ARI_M,BSL_M,LYC_M,SDL_M,CTX_O,MIN_M,RRL_M,BDR_M
Cluster #5	BRG_CG, DUE_U, GFF_CG, SPN_U, ENV_U, PBG_CG

Table 6.17. The code of each stock within each industry plus correlation group for period 4.

Correlation without industry cluster group	Stocks in each cluster
Cluster #1	ALL_CS, AMC_I, SIP_CS, WOW_CS, MTS_CS, CCL_CG, TLS_TC, SKI_U, IFL_F, PTM_F, IAG_F, PPT_F, SUN_F, AIO_I, QBE_F, BEN_F, AMP_F, MQG_F, NAB_F, ANZ_F, WBC_F, CBA_F
Cluster #2	NUF_M, BOQ_F, PRY_H, ASX_F, GUD_CG, IRE_TN, AAD_F
Cluster #3	GNC_CG, FLT_CS, IIN_TN, SAI_I, SGT_TC, DCG_I, SEK_I, LEI_I, WOR_O, HZN_O, ABC_I, IVC_CS, WES_CS, TSE_I, DOW_I, SWM_CS, AHE_CS, UGL_I, BKN_I, ALQ_CG
Cluster #4	AGK_U, MFG_F, TRS_CS, CDD_I, MSB_H, PMV_F, ACR_H, ORG_U, SKE_I, DMP_CS, CPU_I, EWC_U, FWD_CG, ASL_I, SMX_TN, FXL_F, MTU_TC
Cluster #5	SYD_I, SHL_H, RMD_H, TCL_I, NVT_CS, SRX_H, SGN_CS, QUB_I, NZTE_TC, NZSK_CS, NZFB_I, APA_O, SXL_CS, TPI_I, BXB_I

Table 6.18. The code of each stock within each non-industry plus correlation group for period 4.

Period 5 clusters

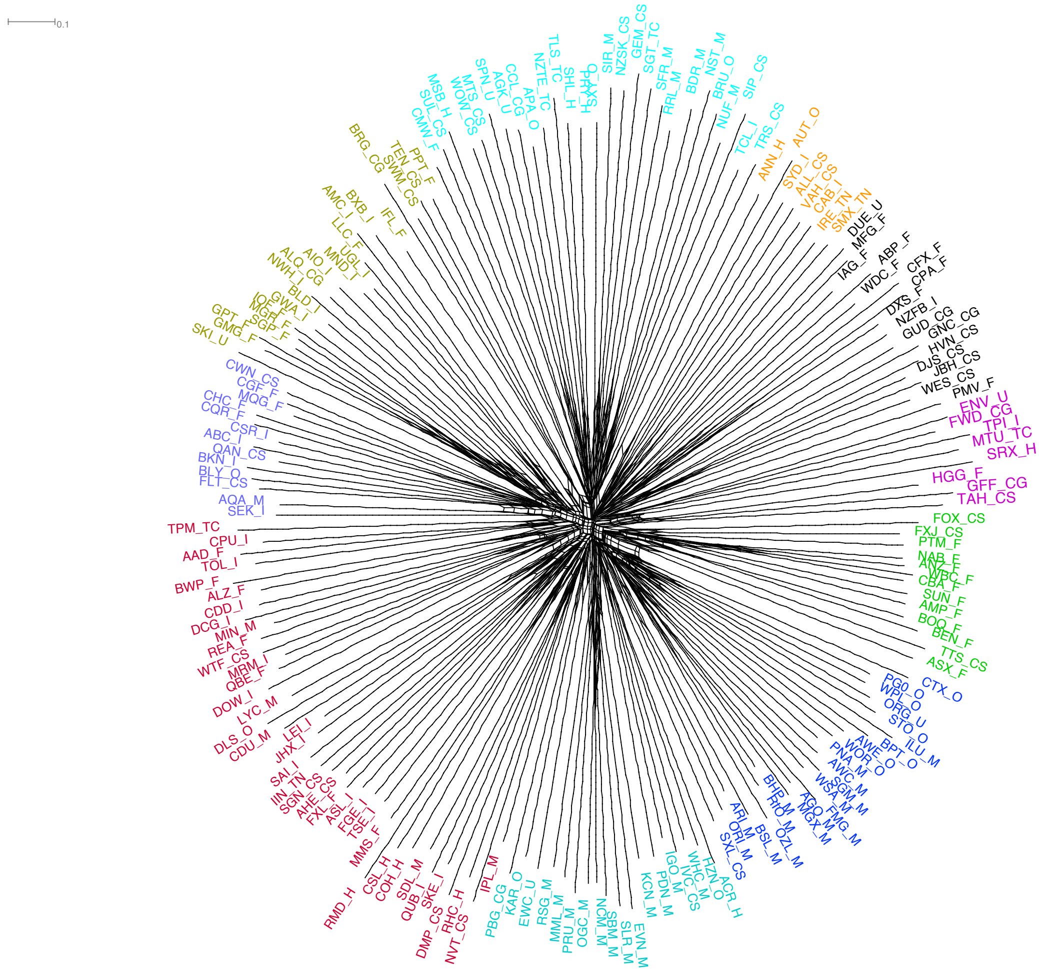


Figure 6.11. ASX 200 185 stocks in period 5: 11 March 2009 to 05 October 2011 is split into 11 correlation clusters suggested by neigh-Net splits graph.

Correlation Cluster Group	Stocks in each correlation cluster group
Cluster #1	ENV_U, FWD_CG, TPI_I, MTU_TC, SRX_H, HGG_F, GFF_CG, TAH_CS
Cluster #2	FOX_CS, FXJ_CS, PTM_F, NAB_F, ANZ_F, WBC_F, CBA_F, SUN_F, AMP_F, BOQ_F, BEN_F, TTS_CS, ASX_F
Cluster #3	CTX_O, PG0_O, WPL_O, ORG_U, STO_O, ILU_M, BPT_O, AWE_O, WOR_O, PNA_M, AWC_M, SGM_M, WSA_M, FMG_M, AGO_M, MGX_M, BHP_M, RIO_M, OZL_M, BSL_M, ARI_M, ORI_M, SXL_CS
Cluster #4	ACR_H, HZN_O, WHC_M, IVC_CS, IGO_M, PDN_M, KCN_M, EVN_M, SLR_M, SBM_M, NCM_M, OGC_M, PRU_M, MML_M, RSG_M, EWC_U, KAR_O, PBG_CG
Cluster #5	IPL_M, NVT_CS, RHC_H, DMP_CS, SKE_I, QUB_I, SDL_M, COH_H, CSL_H, RMD_H, MMS_F, TSE_I, FGE_I, ASL_I, FXL_F, AHE_CS, SGN_CS, IIN_TN, SAI_I, JHX_I, LEI_I, CDU_M, DLS_O, LYC_M, DOW_I, QBE_F, MRM_I, WTF_CS, REA_F, MIN_M, DCG_I, CDD_I, ALZ_F, BWP_F, TOL_I, AAD_F, CPU_I, TPM_TC
Cluster #6	SEK_I, AQA_M, FLT_CS, BLY_O, BKN_I, QAN_CS, ABC_I, CSR_I, CQR_F, CHC_F, MQG_F, CGF_F, CWN_CS
Cluster #7	SKI_U, GMG_F, GPT_F, SGP_F, MGR_F, IOF_F, GWA_I, BLD_I, NWH_I, ALQ_CG, AIO_I, MND_I, UGL_I, LLC_F, AMC_I, BXB_I, IFL_F, BRG_CG, SWM_CS, TEN_CS, PPT_F
Cluster #8	CMW_F, SUL_CS, MSB_H, WOW_CS, MTS_CS, SPN_U, AGK_U, CCL_CG, APA_O, NZTE_TC, TLS_TC, SHL_H, PRY_H, SXY_O, NZSK_CS, SGT_TC, SFR_M, RRL_M, BDR_M, NST_M, BRU_O, NUF_M, SIP_CS, TCL_I, TRS_CS
Cluster #9	ANN_H, AUT_O, SYD_I, ALL_CS, VAH_CS, CAB_I, IRE_TN, SMX_TN
Cluster #10	DUE_U, MFG_F, IAG_F, ABP_F, WDC_F, CFX_F, CPA_F, DXS_F, NZFB_I, GUD_CG, GNC_CG, HVN_CS, DJS_CS, JBH_CS, WES_CS, PMV_F

Table 6.19. The names of stocks in each correlation cluster for period 5: 11 March 2009 to 05 October 2011.

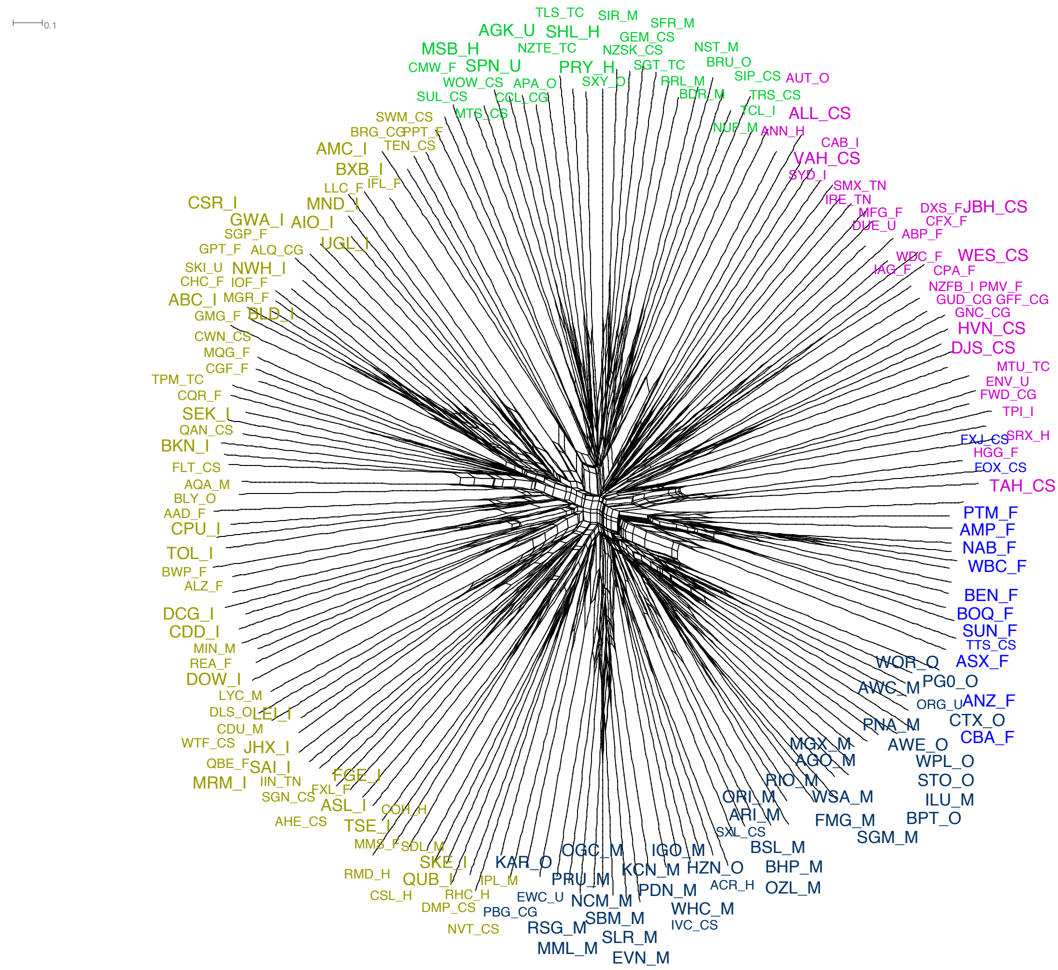


Figure 6.12. The ASX200's 185 stocks in period 5: 11 March 2009 to 05 October 2011 were split into total 10 clusters. Five of the clusters were defined in stock picking strategy #4 (defined by colour and font size) and the other five were defined in the stock picking strategy #5 (defined by colour and size).

Correlation cluster and industry group	Stocks in each cluster
Cluster #1	ALL_CS, VAH_CS, GUD_CG, GNC_CG, HVN_CS, DJS_CS, JBH_CS, WES_CS, GFF_CG, TAH_CS
Cluster #2	PTM_F, NAB_F, ANZ_F, WBC_F, CBA_F, SUN_F, AMP_F, BOQ_F, BEN_F, ASX_F
Cluster #3	CTX_O, PG0_O, WPL_O, STO_O,I LU_M, BPT_O, AWE_O, WOR_O, PNA_M, AWC_M, SGM_M, WSA_M, FMG_M, AGO_M, MGX_M, BHP_M, RIO_M, OZL_M, BSL_M, ARI_M, ORI_M, HZN_O, WHC_M, IGO_M, PDN_M, KCN_M, EVN_M, SLR_M, SBM_M, NCM_M, OGC_M, PRU_M, MML_M, RSG_M, KAR_O
Cluster #4	SKE_I, QUB_I, TSE_I, FGE_I, ASL_I, SAI_I, JHX_I, LEI_I, DOW_I, MRM_I, DCG_I, CDD_I, TOL_I, CPU_I, BKN_I, ABC_I, CSR_I, SEK_I, GWA_I, BLD_I, NWH_I, AIO_I, MND_I, UGL_I, AMC_I, BXB_I
Cluster #5	MSB_H, SPN_U, AGK_U, SHL_H, PRY_H

Table 6.20. The code of each stock within each industry plus correlation group for period 5: 11 March 2009 to 05 October 2011.

Correlation without industry cluster group	**Stocks in each cluster**
Cluster #1	ANN_H, AUT_O, SYD_I, CAB_I, IRE_TN, SMX_TN, DUE_U, MFG_F, IAG_F, ABP_F, WDC_F, CFX_F, CPA_F, DXS_F, NZFB_I, PMV_F, ENV_U, FWD_CG, TPI_I, MTU_TC, SRX_H, HGG_F
Cluster #2	FOX_CS, FXJ_CS, TTS_CS
Cluster #3	ORG_U, SXL_CS, ACR_H, IVC_CS, EWC_U, PBG_CG
Cluster #4	IPL_M, NVT_CS, RHC_H, DMP_CS, SDL_M, COH_H, CSL_H, RMD_H, MMS_F, FXL_F, AHE_CS, SGN_CS, IIN_TN, CDU_M, DLS_O, LYC_M, QBE_F, WTF_CS, REA_F, MIN_M, ALZ_F, BWP_F, AAD_F, TPM_TC, AQA_M, FLT_CS, BLY_O, QAN_CS, CQR_F, CHC_F, MQG_F, CGF_F, CWN_CS, SKI_U, GMG_F, GPT_F, SGP_F, MGR_F, IOF_F, ALQ_CG, LLC_F, IFL_F, BRG_CG, SWM_CS, TEN_CS, PPT_F
Cluster #5	CMW_F, SUL_CS, WOW_CS, MTS_CS, CCL_CG, APA_O, NZTE_TC, TLS_TC, SXY_O, NZSK_CS, SGT_TC, SFR_M, RRL_M, BDR_M, NST_M, BRU_O, NUF_M, SIP_CS, TCL_I, TRS_CS

Table 6.21. The code of each stock within each non-industry plus correlation group for period 5: 11 March 2009 to 05 October 2011.

Simulation Result

Period 2

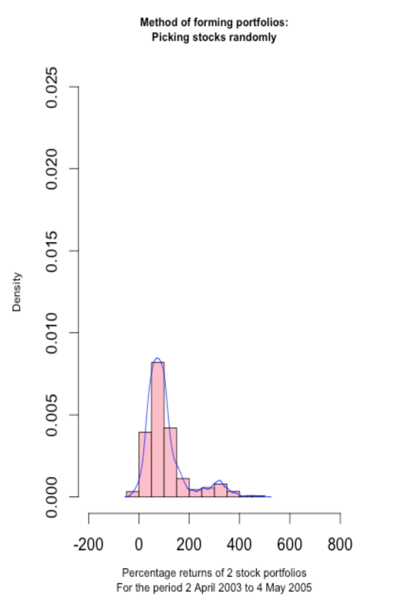


Figure 6.13 (a). The histogram and density curve of the simulated portfolios' percentage returns for period 2: 2 April 2003 to 04 May 2005. Each portfolio contains two stocks which were picked randomly.

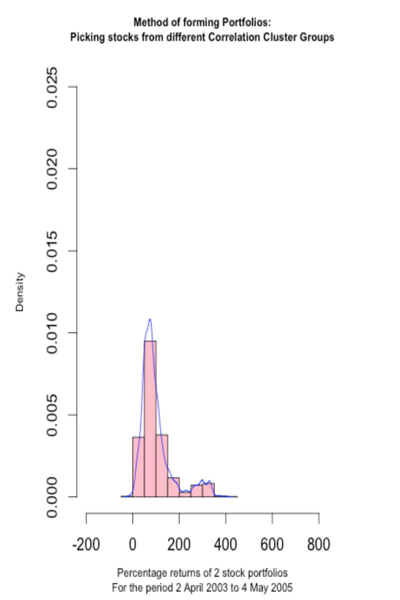


Figure 6.13 (b). The histogram and density curve of the simulated portfolios' percentage returns for period 2: 2 April 2003 to 04 May 2005. Each portfolio contains two stocks which were picked from different correlation clusters suggested by the neighbor-Net split graph produced from the stocks' weekly returns in period 1: 03 May 2000 to 26 March 2003.

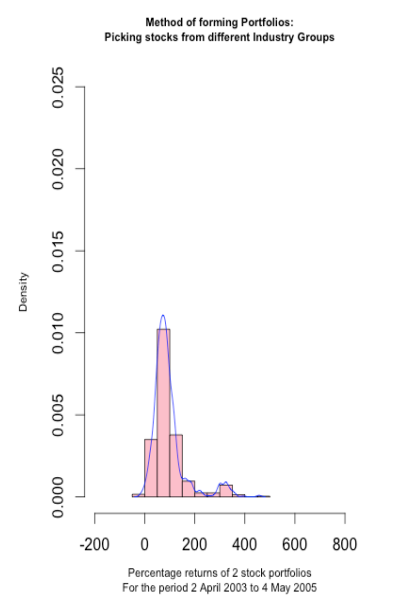


Figure 6.13 (c). The histogram and density curve of the simulated portfolios' percentage returns for period 2: 2 April 2003 to 04 May 2005. Each portfolio contains two stocks which were picked from different industry groups.

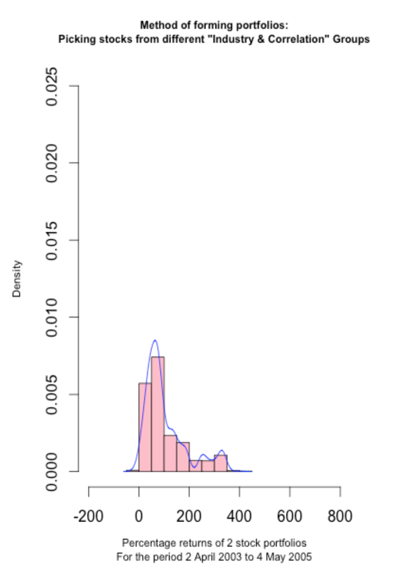


Figure 6.13 (d). The histogram and density curve of the simulated portfolios' percentage returns for period 2: 2 April 2003 to 04 May 2005. Each portfolio contains two stocks which were picked from different clusters that are bounded by both the neighbor-Net split graph produced from the stocks' weekly returns in period 1: 03 May 2000 to 26 March 2003 and one or two industry groups.

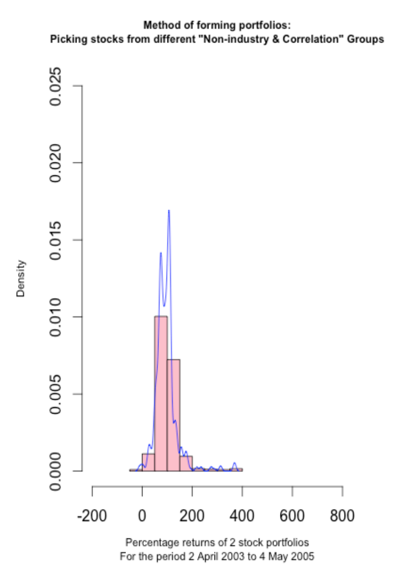
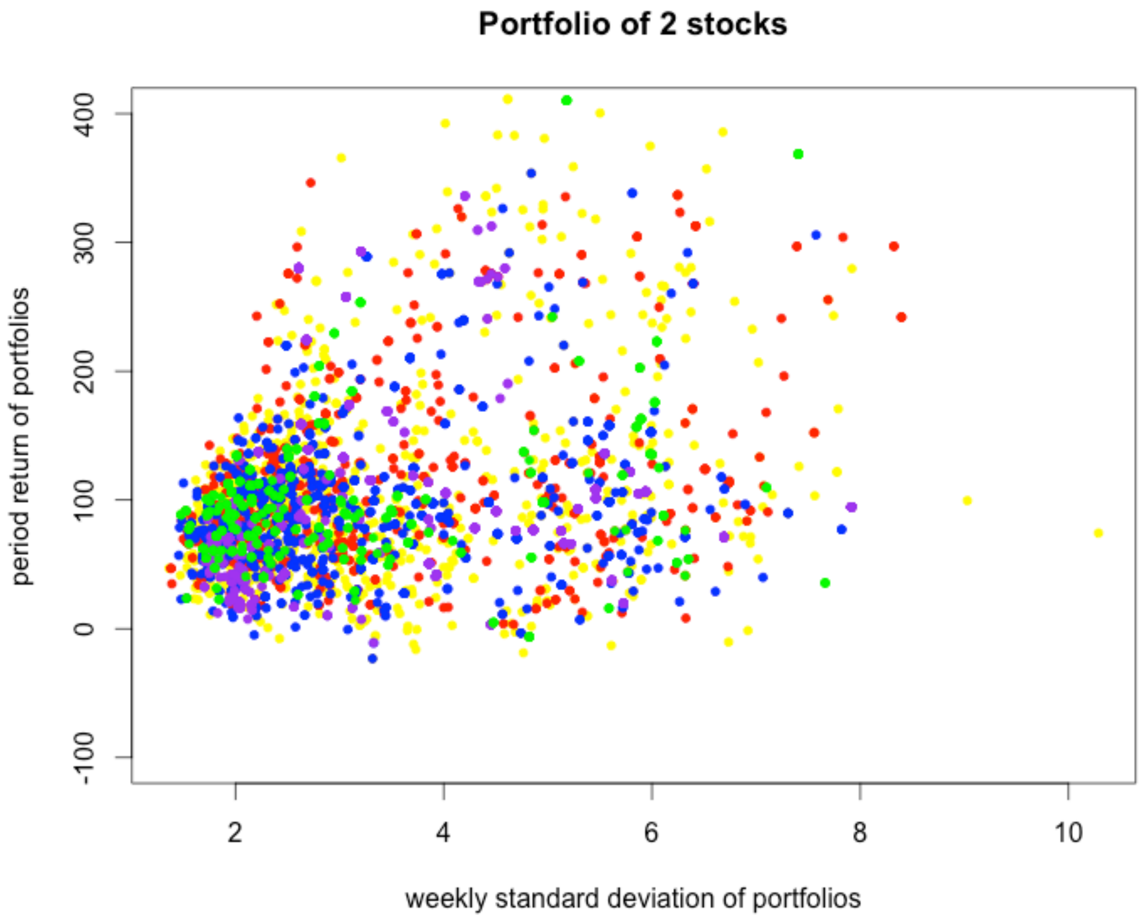


Figure 6.13 (e). The histogram and density curve of the simulated portfolios' percentage returns for period 2: 2 April 2003 to 04 May 2005. Each portfolio contains two stocks which were picked from different clusters that are bounded by the neighbor-Net split graph produced from the stocks' weekly returns in period 1: 03 May 2000 to 26 March 2003 but not industry groups.



- Stocks picked randomly.
- Stocks picked from different correlation clusters suggested by the neighbor-Net splits graph.
- Stocks picked from different industry groups.
- Stocks picked from different correlation and industry groups.
- Stocks picked from different correlation and non-industry groups.

Figure 6.14 Weekly standard deviation and return of each set of simulated 2-stock portfolios for period 2 using the five methods. The period returns and weekly standard deviations are expressed in percentage.

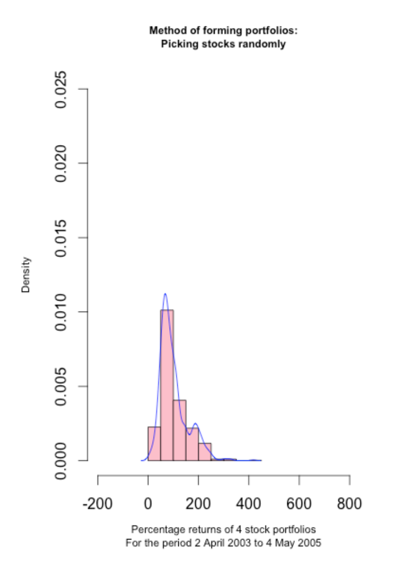


Figure 6.15 (a). The histogram and density curve of the simulated portfolios' percentage returns for period 2: 2 April 2003 to 04 May 2005. Each portfolio contains four stocks which were picked randomly.

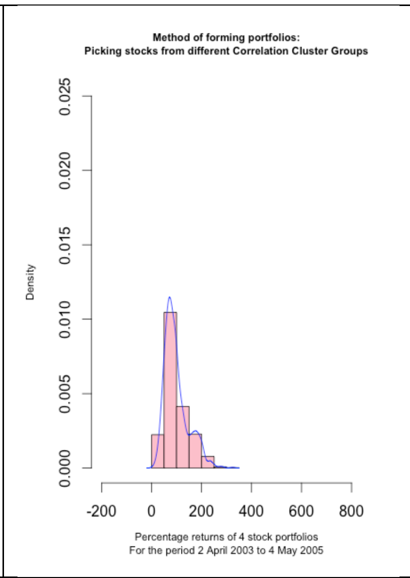


Figure 6.15 (b). The histogram and density curve of the simulated portfolios' percentage returns for period 2: 2 April 2003 to 04 May 2005. Each portfolio contains four stocks which were picked from different correlation clusters suggested by the neighbor-Net split graph produced from the stocks' weekly returns in period 1: 03 May 2000 to 26 March 2003.

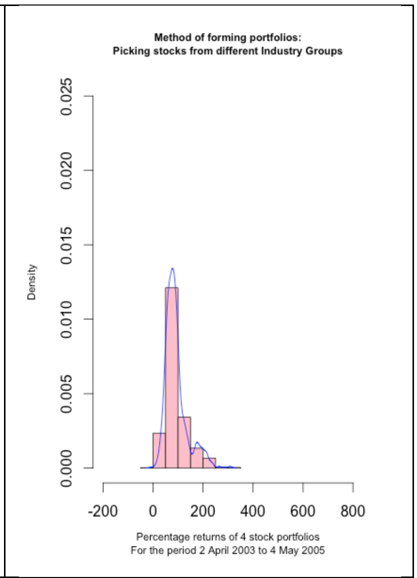


Figure 6.15 (c). The histogram and density curve of the simulated portfolios' percentage returns for period 2: 2 April 2003 to 04 May 2005. Each portfolio contains four stocks which were picked from different industry groups.

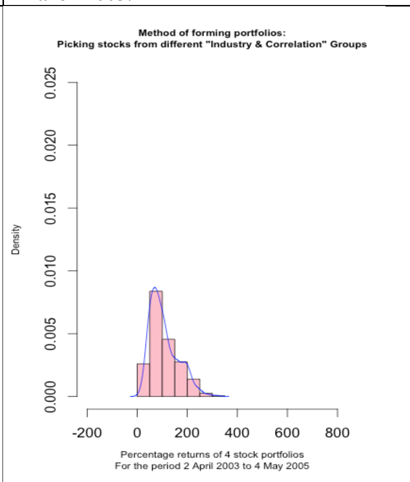


Figure 6.15 (d). The histogram and density curve of the simulated portfolios' percentage returns for period 2: 2 April 2003 to 04 May 2005. Each portfolio contains four stocks which were picked from different clusters that are bounded by both the neighbor-Net split graph produced from the stocks' weekly returns in period 1: 03 May 2000 to 26 March 2003 and one or two industry groups.

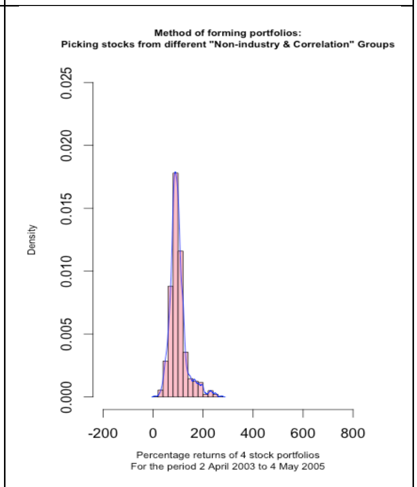
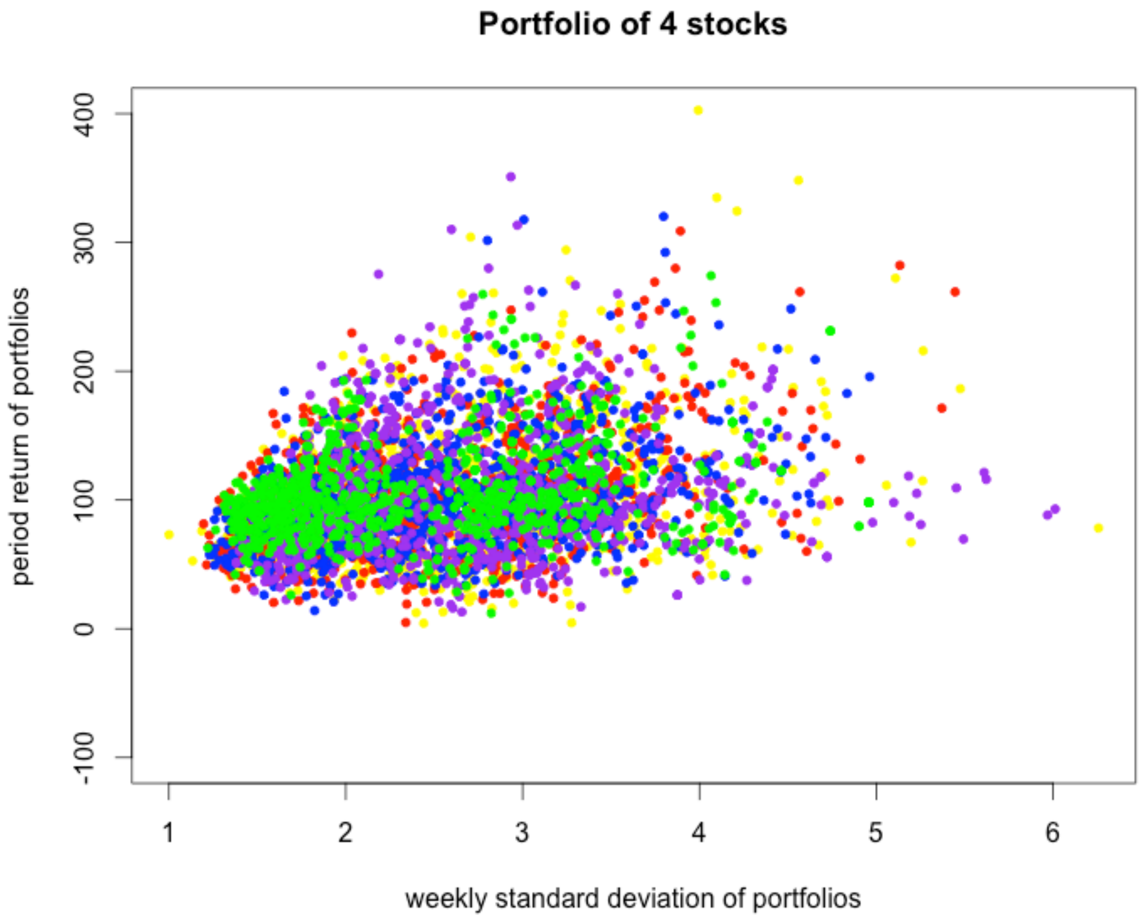


Figure 6.15 (e). The histogram and density curve of the simulated portfolios' percentage returns for period 2: 2 April 2003 to 04 May 2005. Each portfolio contains four stocks which were picked from different clusters that are bounded by the neighbor-Net split graph produced from the stocks' weekly returns in period 1: 03 May 2000 to 26 March 2003 but not industry groups.



- Stocks picked randomly.
- Stocks picked from different correlation clusters suggested by the neighbor-Net splits graph.
- Stocks picked from different industry groups.
- Stocks picked from different correlation and industry groups.
- Stocks picked from different correlation and non-industry groups.

Figure 6.16 Weekly standard deviation and return of each set of simulated 4-stock portfolios for period 2 using the five methods. The period returns and weekly standard deviations are expressed in percentage.

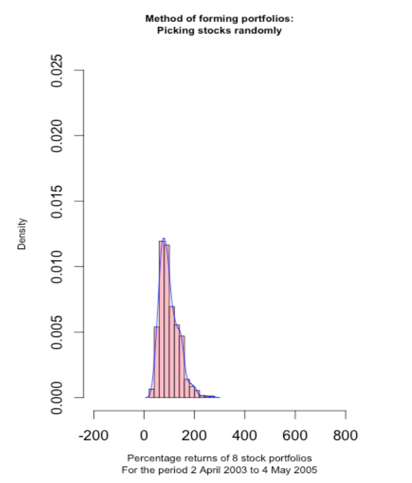


Figure 6.17 (a). The histogram and density curve of the simulated portfolios' percentage returns for period 2: 2 April 2003 to 04 May 2005. Each portfolio contains eight stocks which were picked randomly.

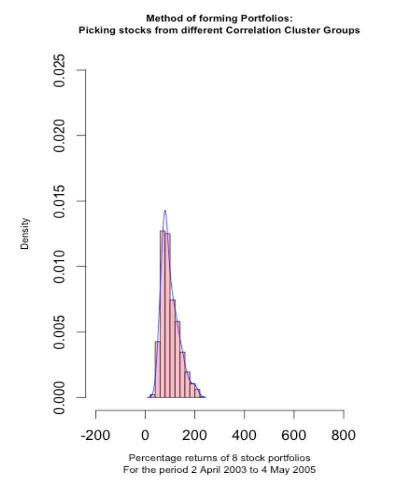


Figure 6.17 (b). The histogram and density curve of the simulated portfolios' percentage returns for period 2: 2 April 2003 to 04 May 2005. Each portfolio contains eight stocks which were picked from different correlation clusters suggested by the neighbor-Net split graph produced from the stocks' weekly returns in period 1: 03 May 2000 to 26 March 2003.

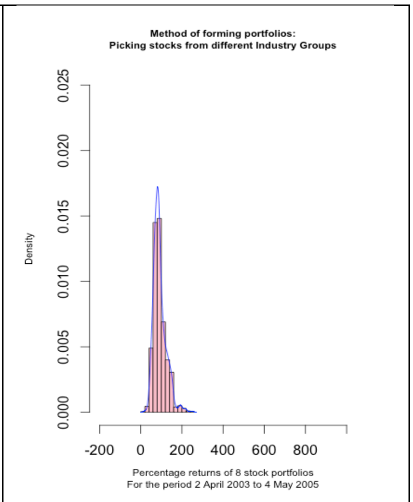


Figure 6.17 (c). The histogram and density curve of the simulated portfolios' percentage returns for period 2: 2 April 2003 to 04 May 2005. Each portfolio contains eight stocks which were picked from different industry groups.

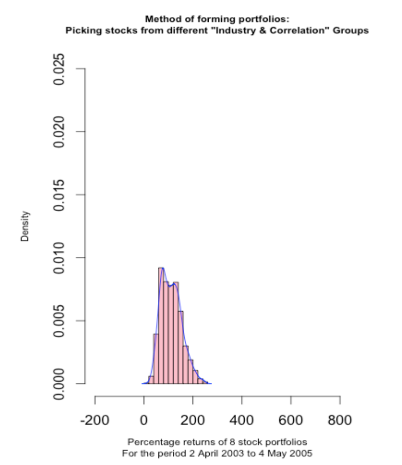


Figure 6.17 (d). The histogram and density curve of the simulated portfolios' percentage returns for period 2: 2 April 2003 to 04 May 2005. Each portfolio contains eight stocks which were picked from different clusters that are bounded by both the neighbor-Net split graph produced from the stocks' weekly returns in period 1: 03 May 2000 to 26 March 2003 and one or two industry groups.

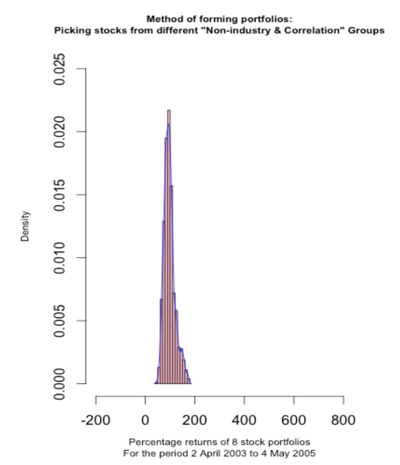
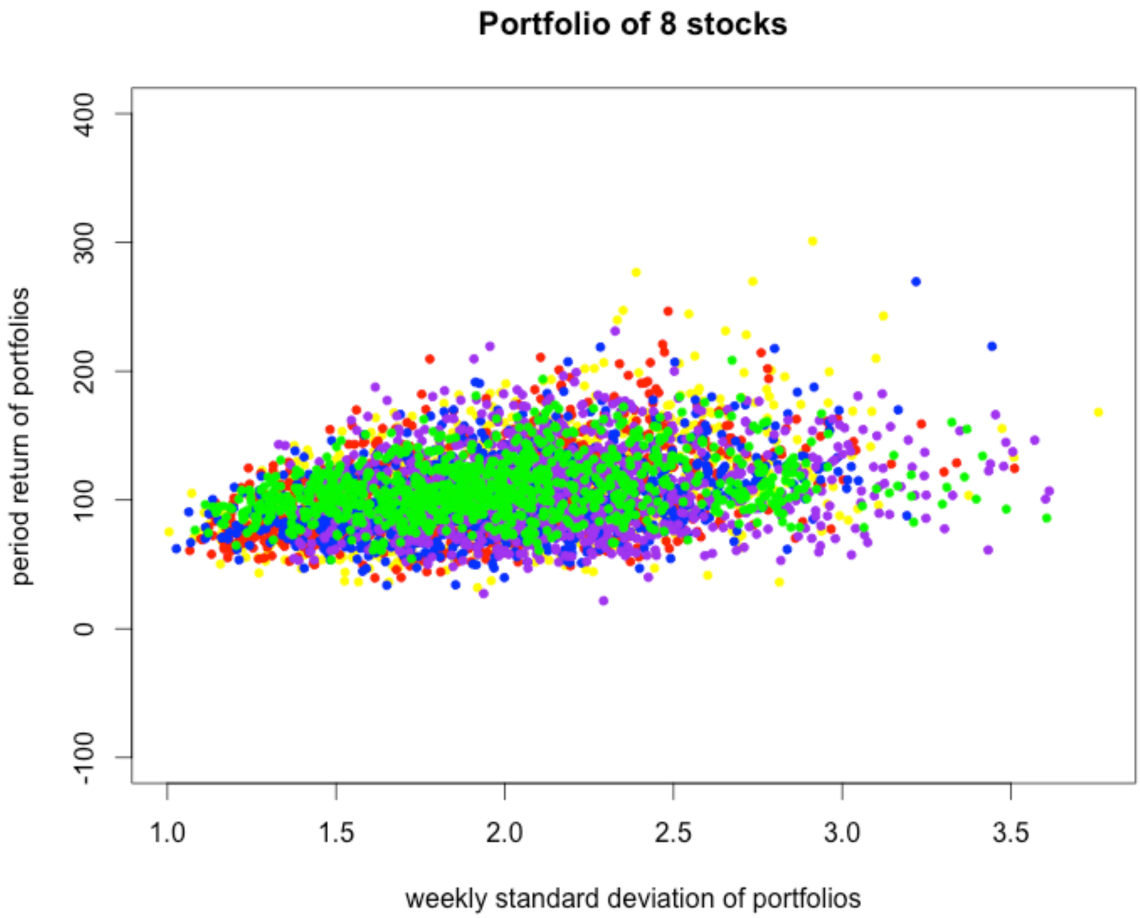


Figure 6.17 (e). The histogram and density curve of the simulated portfolios' percentage returns for period 2: 2 April 2003 to 04 May 2005. Each portfolio contains eight stocks which were picked from different clusters that are bounded by the neighbor-Net split graph produced from the stocks' weekly returns in period 1: 03 May 2000 to 26 March 2003 but not industry groups.



- Stocks picked randomly.
- Stocks picked from different correlation clusters suggested by the neighbor-Net splits graph.
- Stocks picked from different industry groups.
- Stocks picked from different correlation and industry groups.
- Stocks picked from different correlation and non-industry groups.

Figure 6.18 Weekly standard deviation and return of each set of simulated 8-stock portfolios for period 2 using the five methods. The period returns and weekly standard deviations are expressed in percentage.

Period 3

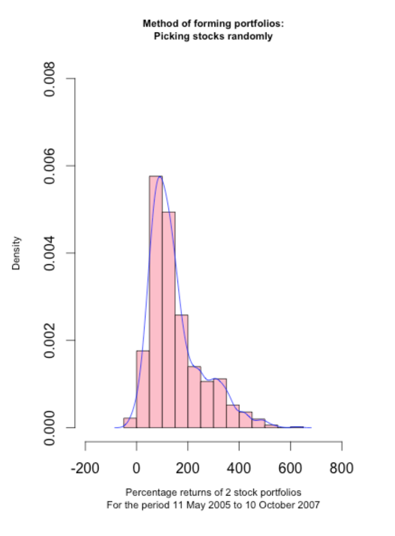


Figure 6.19 (a). The histogram and density curve of the simulated portfolios' percentage returns for period 3: 11 May 2005 to 10 October 2007. Each portfolio contains two stocks which were picked randomly.

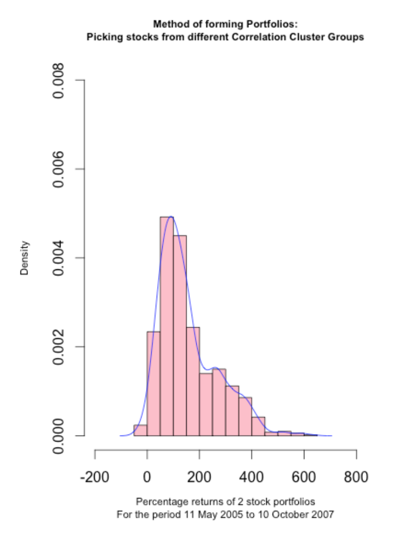


Figure 6.19 (b). The histogram and density curve of the simulated portfolios' percentage returns for period 3: 11 May 2005 to 10 October 2007. Each portfolio contains two stocks which were picked from different correlation clusters suggested by the neighbor-Net split graph produced from the stocks' weekly returns in period 2: 02 April 2003 to 04 May 2005.

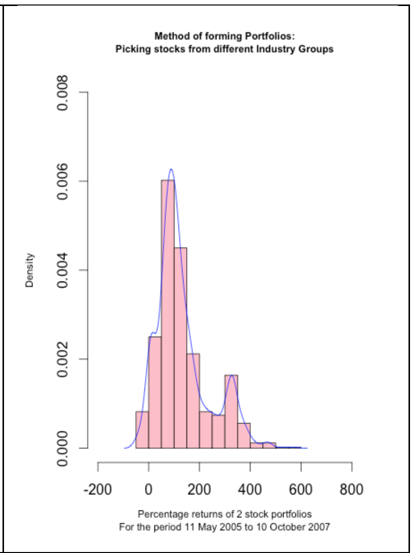


Figure 6.19 (c). The histogram and density curve of the simulated portfolios' percentage returns for period 3: 11 May 2005 to 10 October 2007. Each portfolio contains two stocks which were picked from different industry groups.

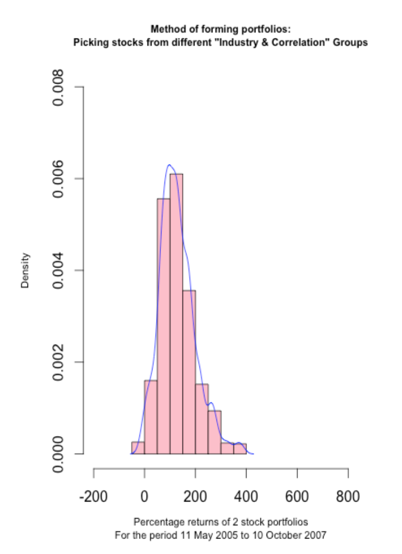


Figure 6.19 (d). The histogram and density curve of the simulated portfolios' percentage returns for period 3: 11 May 2005 to 10 October 2007. Each portfolio contains two stocks which were picked from different clusters that are bounded by both the neighbor-Net split graph produced from the stocks' weekly returns in period 2: 02 April 2003 to 04 May 2005.

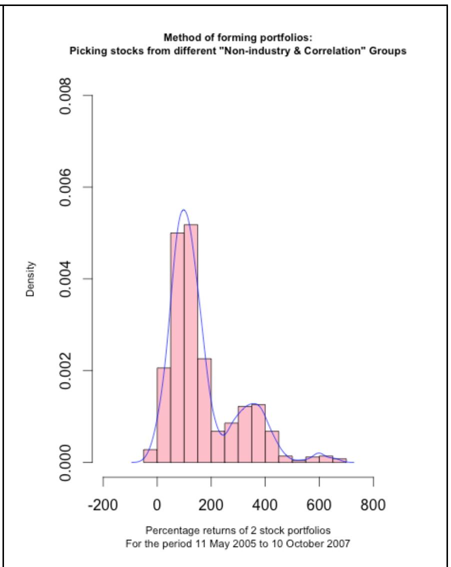
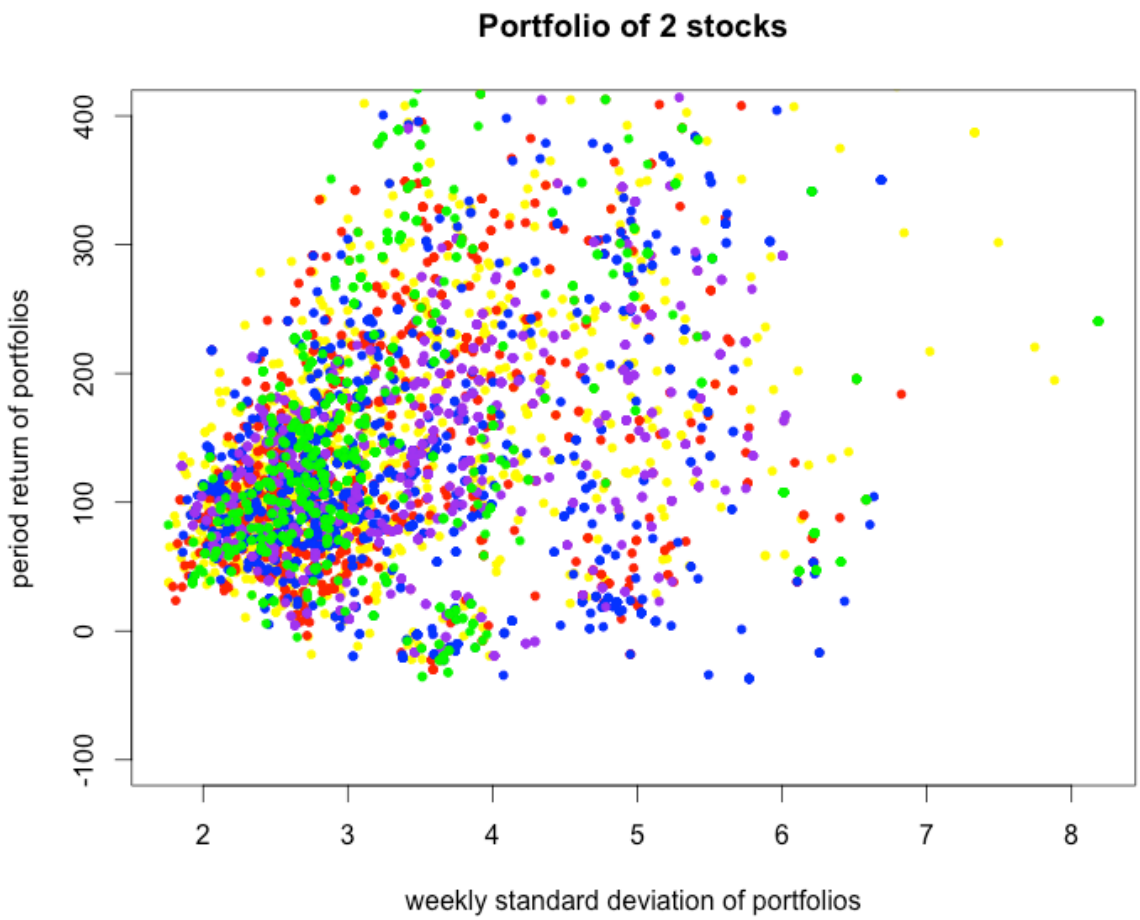


Figure 6.19 (e). The histogram and density curve of the simulated portfolios' percentage returns for period 3: 11 May 2005 to 10 October 2007. Each portfolio contains two stocks which were picked from different clusters that are bounded by the neighbor-Net split graph produced from the stocks' weekly returns in period 2: 02 April 2003 to 04 May 2005 but not industry groups.



- Stocks picked randomly.
- Stocks picked from different correlation clusters suggested by the neighbor-Net splits graph.
- Stocks picked from different industry groups.
- Stocks picked from different correlation and industry groups.
- Stocks picked from different correlation and non-industry groups.

Figure 6.20 Weekly standard deviation and return of each set of simulated 2-stock portfolios for period 3 using the five methods. The period returns and weekly standard deviations are expressed in percentage.

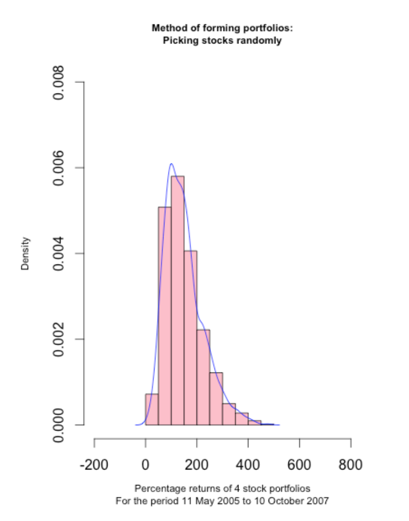


Figure 6.21 (a). The histogram and density curve of the simulated portfolios' percentage returns for period 3: 11 May 2005 to 10 October 2007. Each portfolio contains four stocks which were picked randomly.

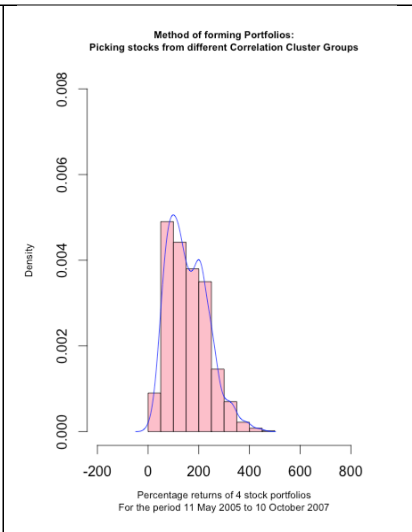


Figure 6.21 (b). The histogram and density curve of the simulated portfolios' percentage returns for period 3: 11 May 2005 to 10 October 2007. Each portfolio contains four stocks which were picked from different correlation clusters suggested by the neighbor-Net split graph produced from the stocks' weekly returns in period 2: 02 April 2003 to 04 May 2005.

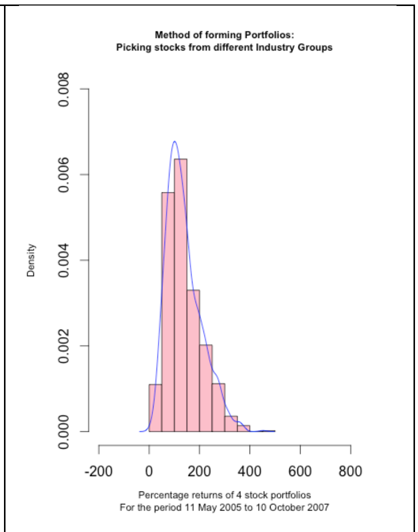


Figure 6.21 (c). The histogram and density curve of the simulated portfolios' percentage returns for period 3: 11 May 2005 to 10 October 2007. Each portfolio contains four stocks which were picked from different industry groups.

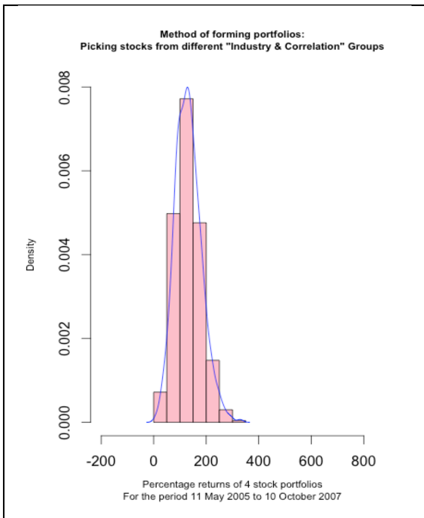


Figure 6.21 (d). The histogram and density curve of the simulated portfolios' percentage returns for period 3: 11 May 2005 to 10 October 2007. Each portfolio contains four stocks which were picked from different clusters that are bounded by both the neighbor-Net split graph produced from the stocks' weekly returns in period 2: 02 April 2003 to 04 May 2005.

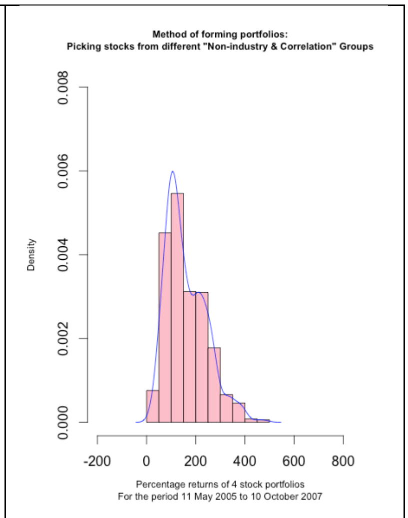


Figure 6.21 (e). The histogram and density curve of the simulated portfolios' percentage returns for period 3: 11 May 2005 to 10 October 2007. Each portfolio contains four stocks which were picked from different clusters that are bounded by the neighbor-Net split graph produced from the stocks' weekly returns in period 2: 02 April 2003 to 04 May 2005 but not industry groups.

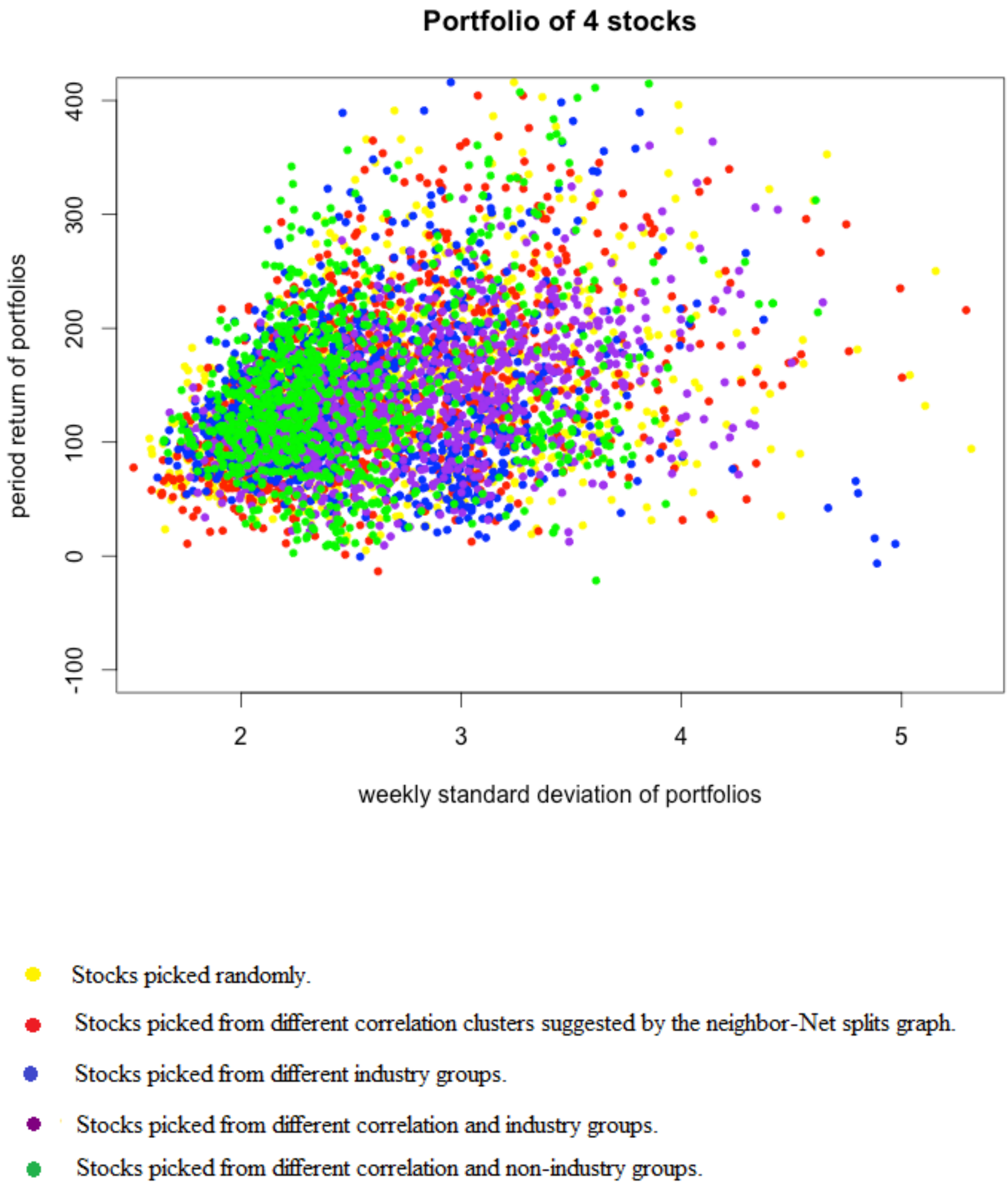


Figure 6.22 Weekly standard deviation and return of each set of simulated 4-stock portfolios for period 3 using the five methods. The period returns and weekly standard deviations are expressed in percentage.

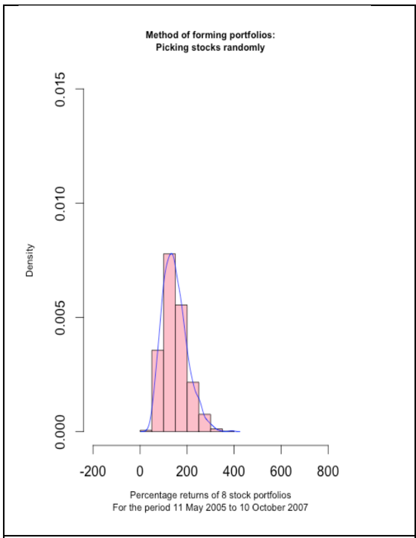


Figure 6.23 (a). The histogram and density curve of the simulated portfolios' percentage returns for period 3: 11 May 2005 to 10 October 2007. Each portfolio contains eight stocks which were picked randomly.

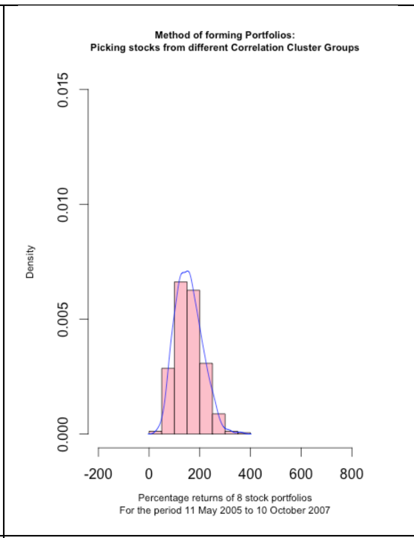


Figure 6.23 (b). The histogram and density curve of the simulated portfolios' percentage returns for period 3: 11 May 2005 to 10 October 2007. Each portfolio contains eight stocks which were picked from different correlation clusters suggested by the neighbor-Net split graph produced from the stocks' weekly returns in period 2: 02 April 2003 to 04 May 2005.

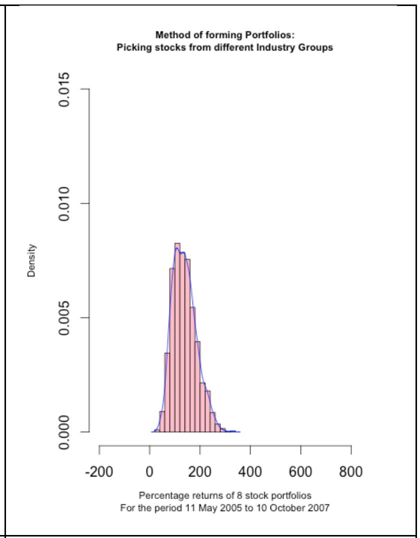


Figure 6.23 (c). The histogram and density curve of the simulated portfolios' percentage returns for period 3: 11 May 2005 to 10 October 2007. Each portfolio contains eight stocks which were picked from different industry groups.

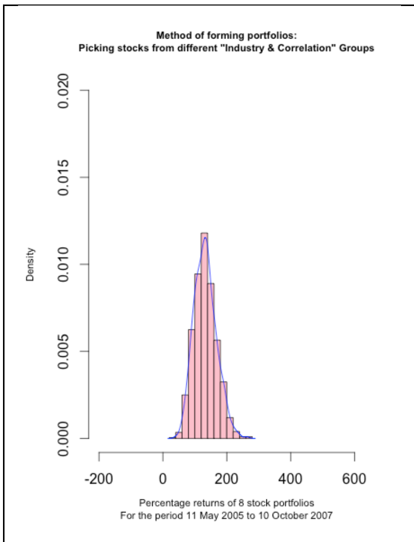


Figure 6.23 (d). The histogram and density curve of the simulated portfolios' percentage returns for period 3: 11 May 2005 to 10 October 2007. Each portfolio contains eight stocks which were picked from different clusters that are bounded by both the neighbor-Net split graph produced from the stocks' weekly returns in period 2: 02 April 2003 to 04 May 2005.

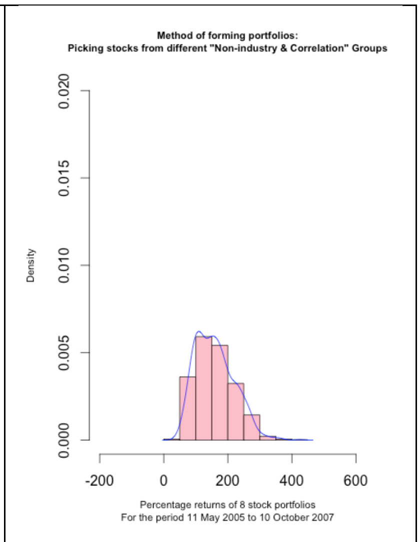
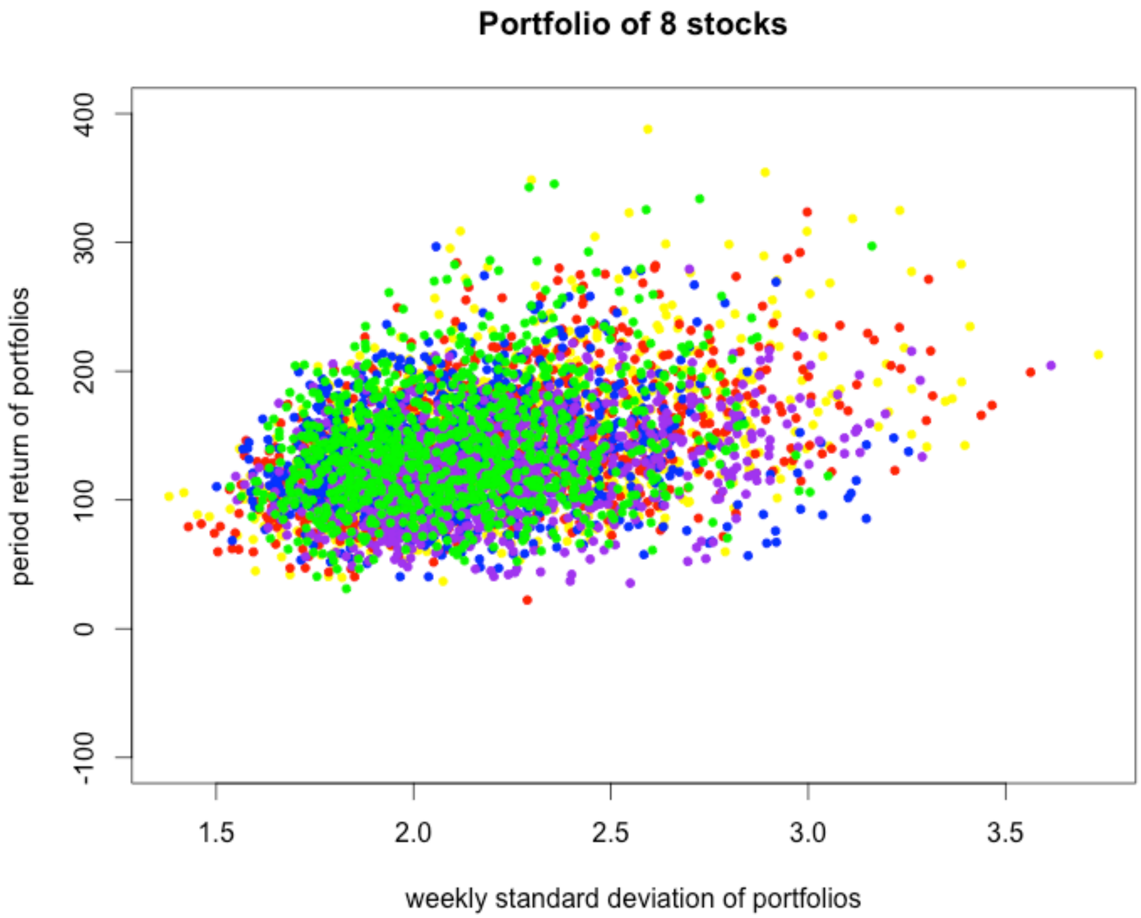


Figure 6.23 (e). The histogram and density curve of the simulated portfolios' percentage returns for period 3: 11 May 2005 to 10 October 2007. Each portfolio contains eight stocks which were picked from different clusters that are bounded by the neighbor-Net split graph produced from the stocks' weekly returns in period 2: 02 April 2003 to 04 May 2005 but not industry groups.



- Stocks picked randomly.
- Stocks picked from different correlation clusters suggested by the neighbor-Net splits graph.
- Stocks picked from different industry groups.
- Stocks picked from different correlation and industry groups.
- Stocks picked from different correlation and non-industry groups.

Figure 6.24 Weekly standard deviation and return of each set of simulated 8-stock portfolios for period 3 using the five methods. The period returns and weekly standard deviations are expressed in percentage.

Period 4

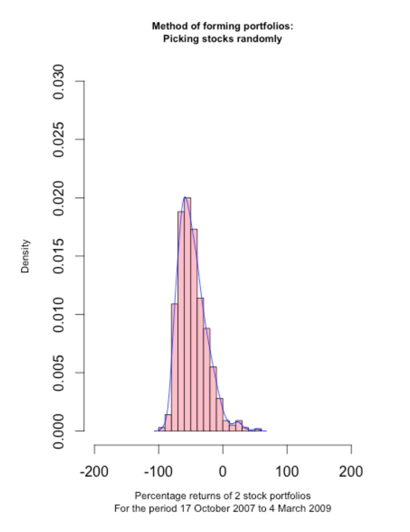


Figure 6.25 (a). The histogram and density curve of the simulated portfolios' percentage returns for period 4: 17 October 2007 to 4 March 2009. Each portfolio contains two stocks which were picked randomly.

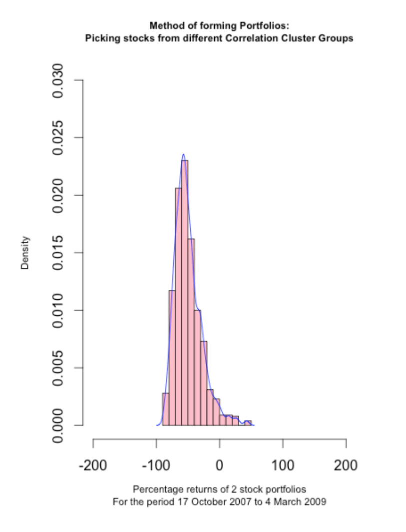


Figure 6.25 (b). The histogram and density curve of the simulated portfolios' percentage returns for period 4: 17 October 2007 to 4 March 2009. Each portfolio contains two stocks which were picked from different correlation clusters suggested by the neighbor-Net split graph produced from the stocks' weekly returns in period 3: 11 May 2005 to 10 October 2007.

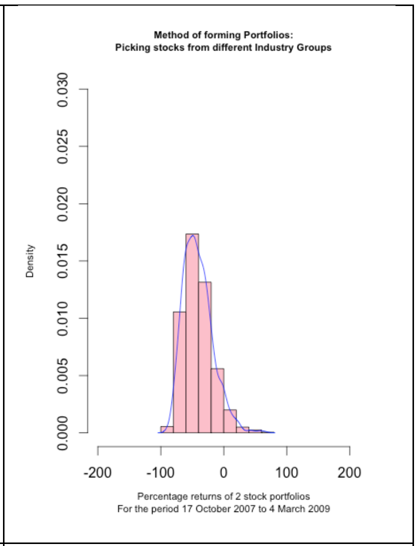


Figure 6.25 (c). The histogram and density curve of the simulated portfolios' percentage returns for period 4: 17 October 2007 to 4 March 2009. Each portfolio contains two stocks which were picked from different industry groups.

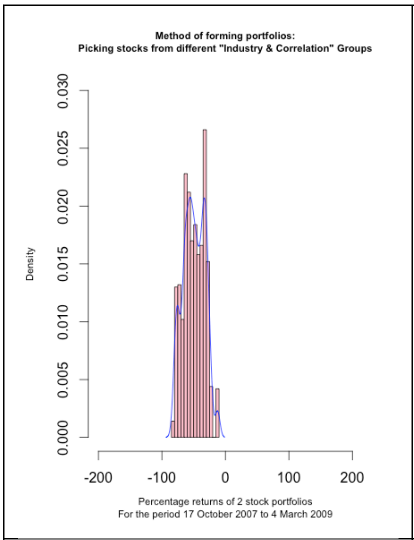


Figure 6.25 (d). The histogram and density curve of the simulated portfolios' percentage returns for period 4: 17 October 2007 to 4 March 2009. Each portfolio contains two stocks which were picked from different clusters that are bounded by both the neighbor-Net split graph produced from the stocks' weekly returns in period 3: 11 May 2005 to 10 October 2007.

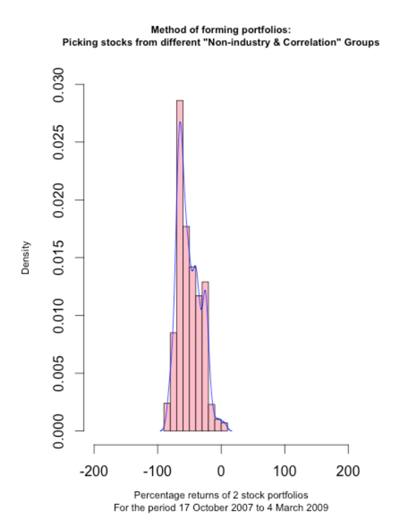


Figure 6.25 (e). The histogram and density curve of the simulated portfolios' percentage returns for period 4: 17 October 2007 to 4 March 2009. Each portfolio contains two stocks which were picked from different clusters that are bounded by the neighbor-Net split graph produced from the stocks' weekly returns in period 3: 11 May 2005 to 10 October 2007 but not industry groups.

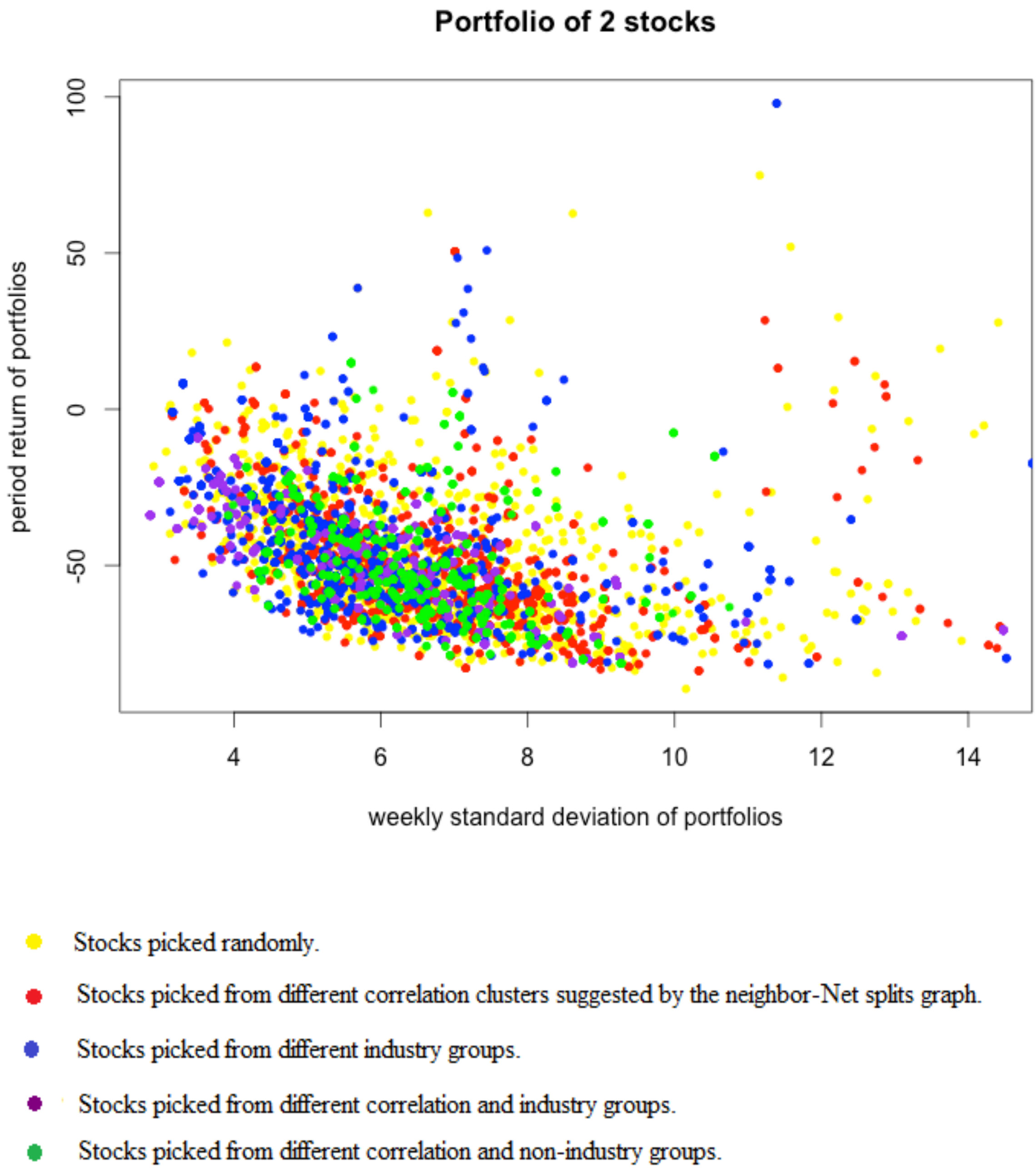


Figure 6.26 Weekly standard deviation and return of each set of simulated 2-stock portfolios for period 4 using the five methods. The period returns and weekly standard deviations are expressed in percentage.

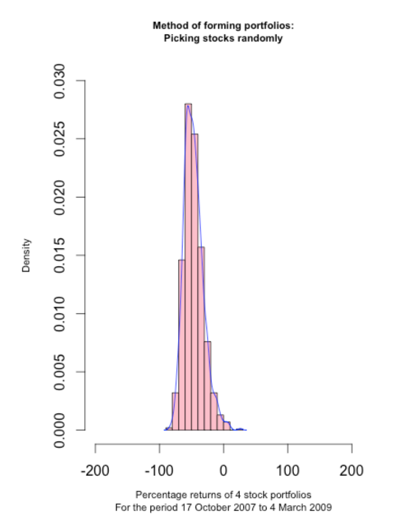


Figure 6.27 (a). The histogram and density curve of the simulated portfolios' percentage returns for period 4: 17 October 2007 to 4 March 2009. Each portfolio contains four stocks which were picked randomly.

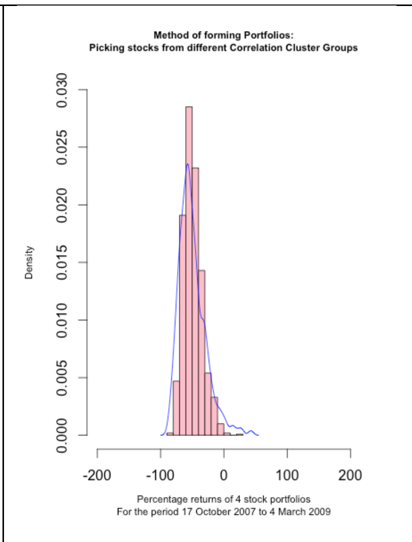


Figure 6.27 (b). The histogram and density curve of the simulated portfolios' percentage returns for period 4: 17 October 2007 to 4 March 2009. Each portfolio contains four stocks which were picked from different correlation clusters suggested by the neighbor-Net split graph produced from the stocks' weekly returns in period 3: 11 May 2005 to 10 October 2007.

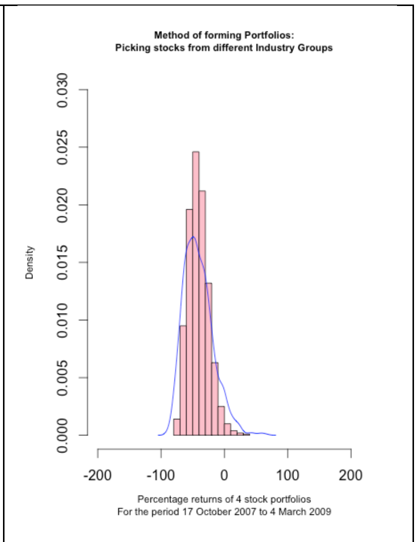


Figure 6.27 (c). The histogram and density curve of the simulated portfolios' percentage returns for period 4: 17 October 2007 to 4 March 2009. Each portfolio contains four stocks which were picked from different industry groups.

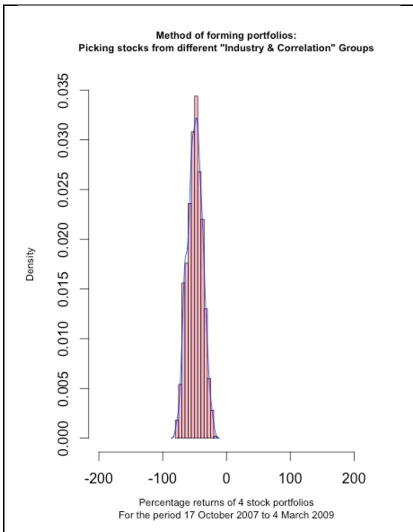


Figure 6.27 (d). The histogram and density curve of the simulated portfolios' percentage returns for period 4: 17 October 2007 to 4 March 2009. Each portfolio contains four stocks which were picked from different clusters that are bounded by both the neighbor-Net split graph produced from the stocks' weekly returns in period 3: 11 May 2005 to 10 October 2007.

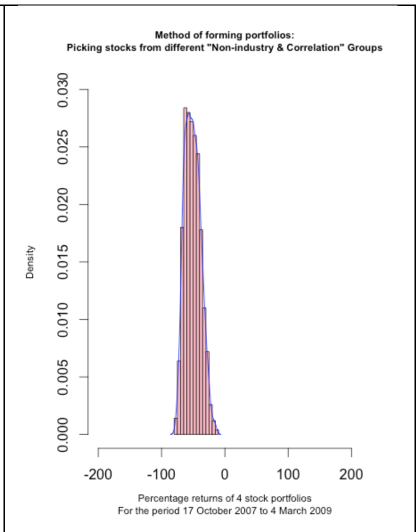
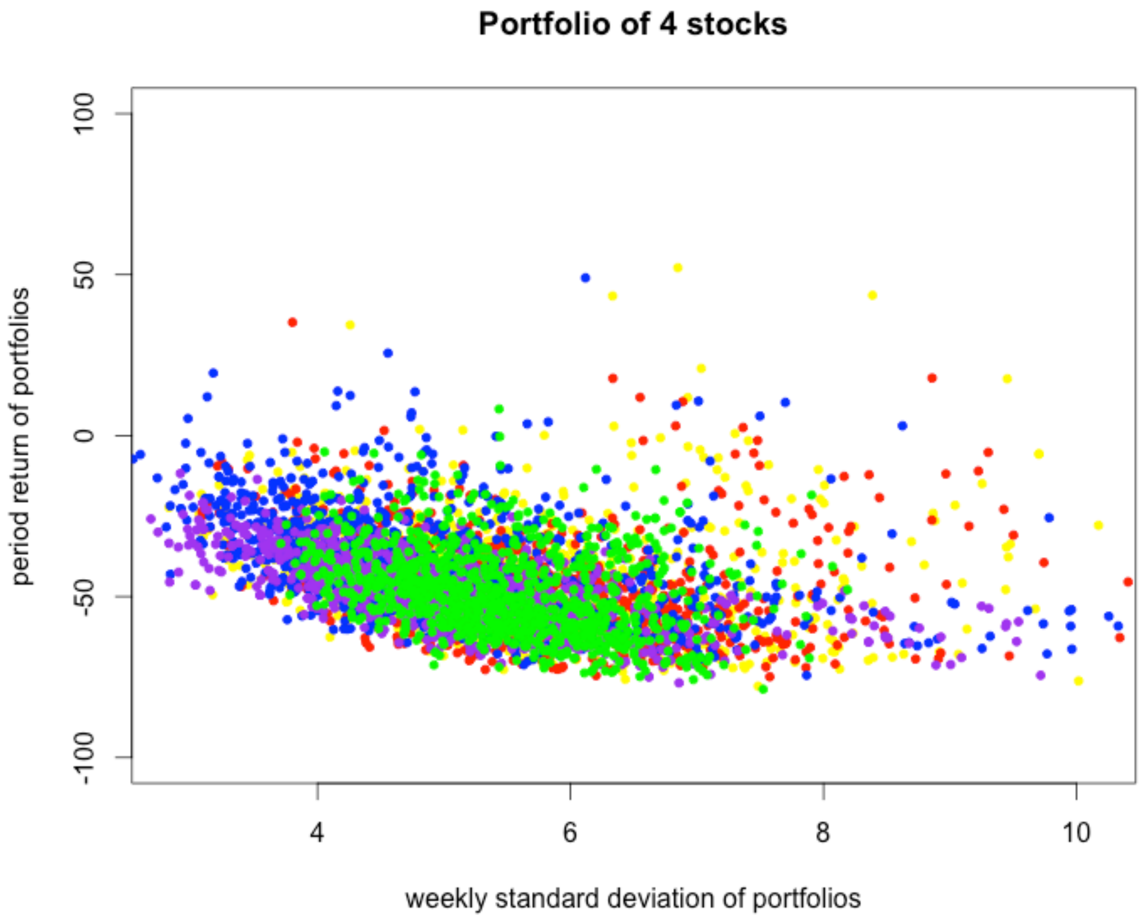


Figure 6.27 (e). The histogram and density curve of the simulated portfolios' percentage returns for period 4: 17 October 2007 to 4 March 2009. Each portfolio contains four stocks which were picked from different clusters that are bounded by the neighbor-Net split graph produced from the stocks' weekly returns in period 3: 11 May 2005 to 10 October 2007 but not industry groups.



- Stocks picked randomly.
- Stocks picked from different correlation clusters suggested by the neighbor-Net splits graph.
- Stocks picked from different industry groups.
- Stocks picked from different correlation and industry groups.
- Stocks picked from different correlation and non-industry groups.

Figure 6.28 Weekly standard deviation and return of each set of simulated 4-stock portfolios for period 4 using the five methods. The period returns and weekly standard deviations are expressed in percentage.

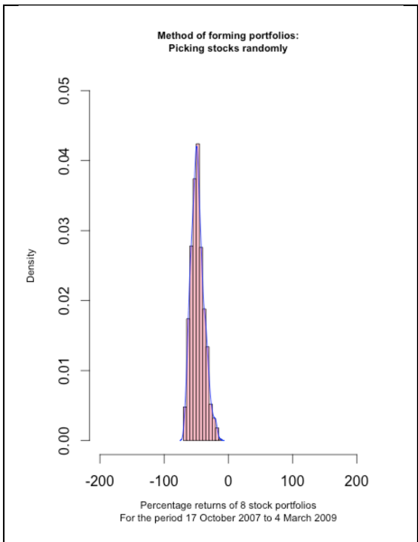


Figure 6.29 (a). The histogram and density curve of the simulated portfolios' percentage returns for period 4: 17 October 2007 to 4 March 2009. Each portfolio contains eight stocks which were picked randomly.

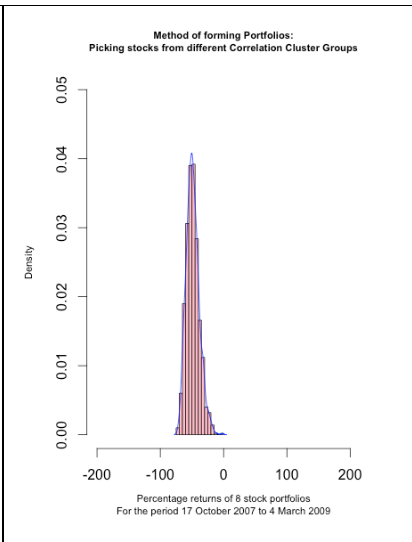


Figure 6.29 (b). The histogram and density curve of the simulated portfolios' percentage returns for period 4: 17 October 2007 to 4 March 2009. Each portfolio contains eight stocks which were picked from different correlation clusters suggested by the neighbor-Net split graph produced from the stocks' weekly returns in period 3: 11 May 2005 to 10 October 2007.

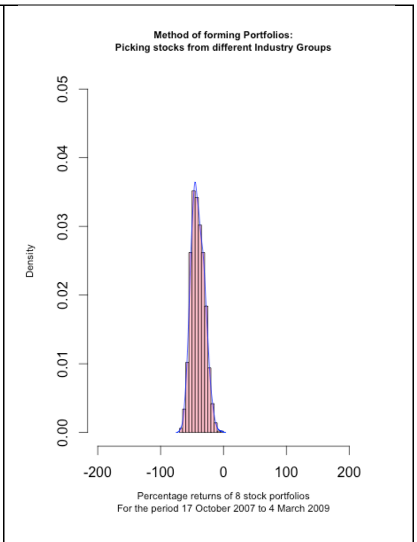


Figure 6.29 (c). The histogram and density curve of the simulated portfolios' percentage returns for period 4: 17 October 2007 to 4 March 2009. Each portfolio contains eight stocks which were picked from different industry groups.

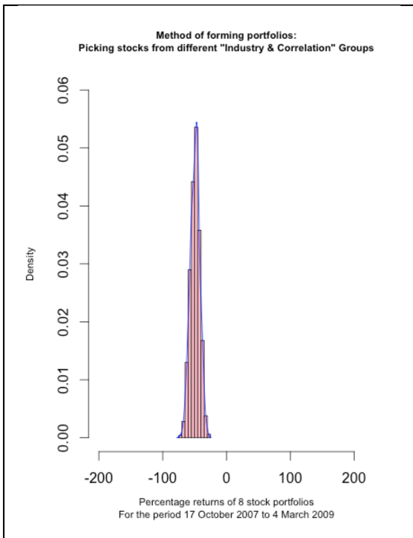


Figure 6.29 (d). The histogram and density curve of the simulated portfolios' percentage returns for period 4: 17 October 2007 to 4 March 2009. Each portfolio contains eight stocks which were picked from different clusters that are bounded by both the neighbor-Net split graph produced from the stocks' weekly returns in period 3: 11 May 2005 to 10 October 2007.

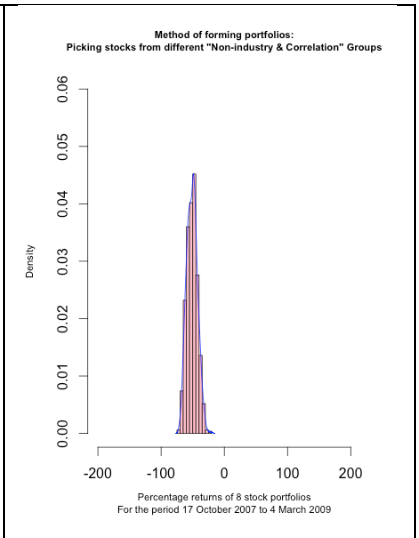
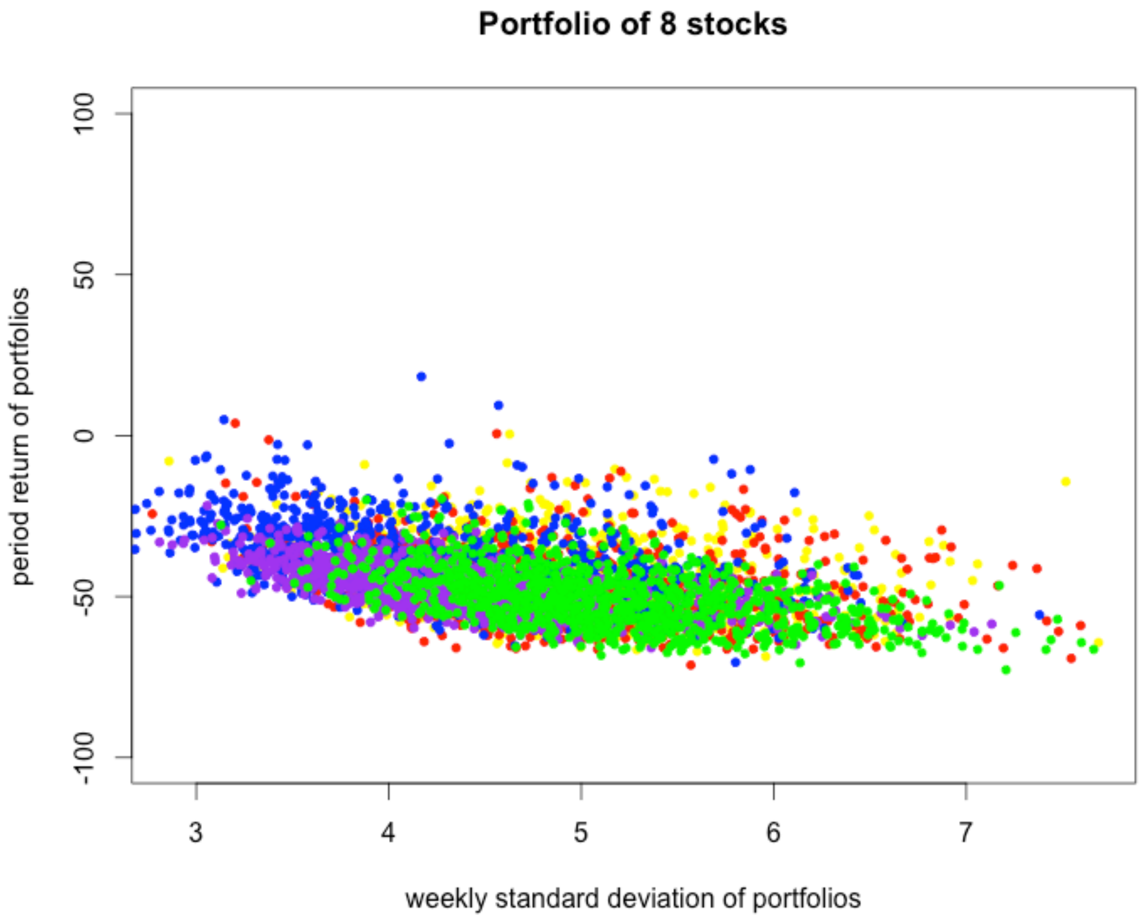


Figure 6.29 (e). The histogram and density curve of the simulated portfolios' percentage returns for period 4: 17 October 2007 to 4 March 2009. Each portfolio contains eight stocks which were picked from different clusters that are bounded by the neighbor-Net split graph produced from the stocks' weekly returns in period 3: 11 May 2005 to 10 October 2007 but not industry groups.



- Stocks picked randomly.
- Stocks picked from different correlation clusters suggested by the neighbor-Net splits graph.
- Stocks picked from different industry groups.
- Stocks picked from different correlation and industry groups.
- Stocks picked from different correlation and non-industry groups.

Figure 6.30 Weekly standard deviation and return of each set of simulated 8-stock portfolios for period 4 using the five methods. The period returns and weekly standard deviations are expressed in percentage.

Period 5

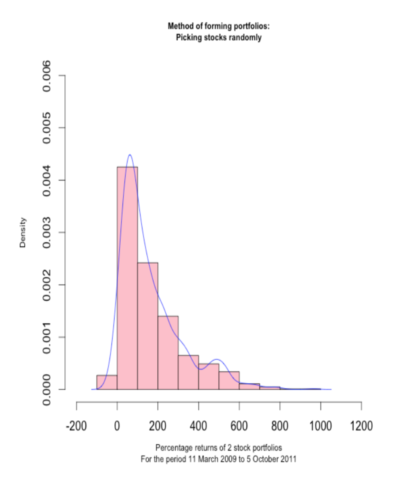


Figure 6.31 (a). The histogram and density curve of the simulated portfolios' percentage returns for period 5: 11 March 2009 to 5 October 2011. Each portfolio contains two stocks which were picked randomly.

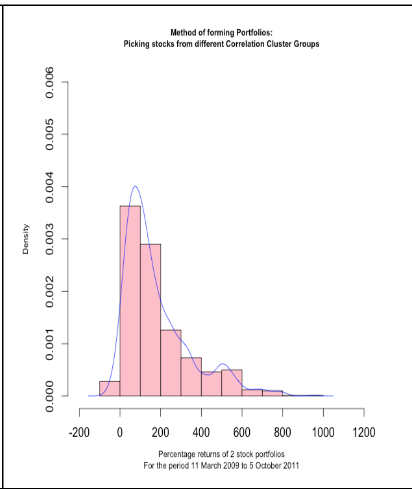


Figure 6.31 (b). The histogram and density curve of the simulated portfolios' percentage returns for period 5: 11 March 2009 to 5 October 2011. Each portfolio contains two stocks which were picked from different correlation clusters suggested by the neighbor-Net split graph produced from the stocks' weekly returns in period 4: 17 October 2007 to 4 March 2009.

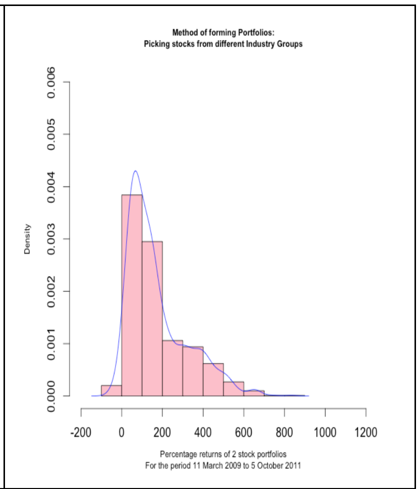


Figure 6.31 (c). The histogram and density curve of the simulated portfolios' percentage returns for period 5: 11 March 2009 to 5 October 2011. Each portfolio contains two stocks which were picked from different industry groups.

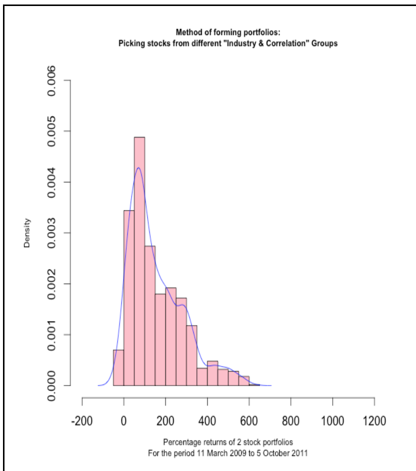


Figure 6.31 (d). The histogram and density curve of the simulated portfolios' percentage returns for period 5: 11 March 2009 to 5 October 2011. Each portfolio contains two stocks which were picked from different clusters that are bounded by both the neighbor-Net split graph produced from the stocks' weekly returns in period 4: 17 October 2007 to 4 March 2009.

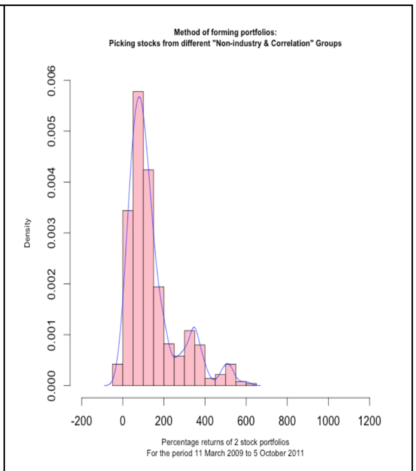
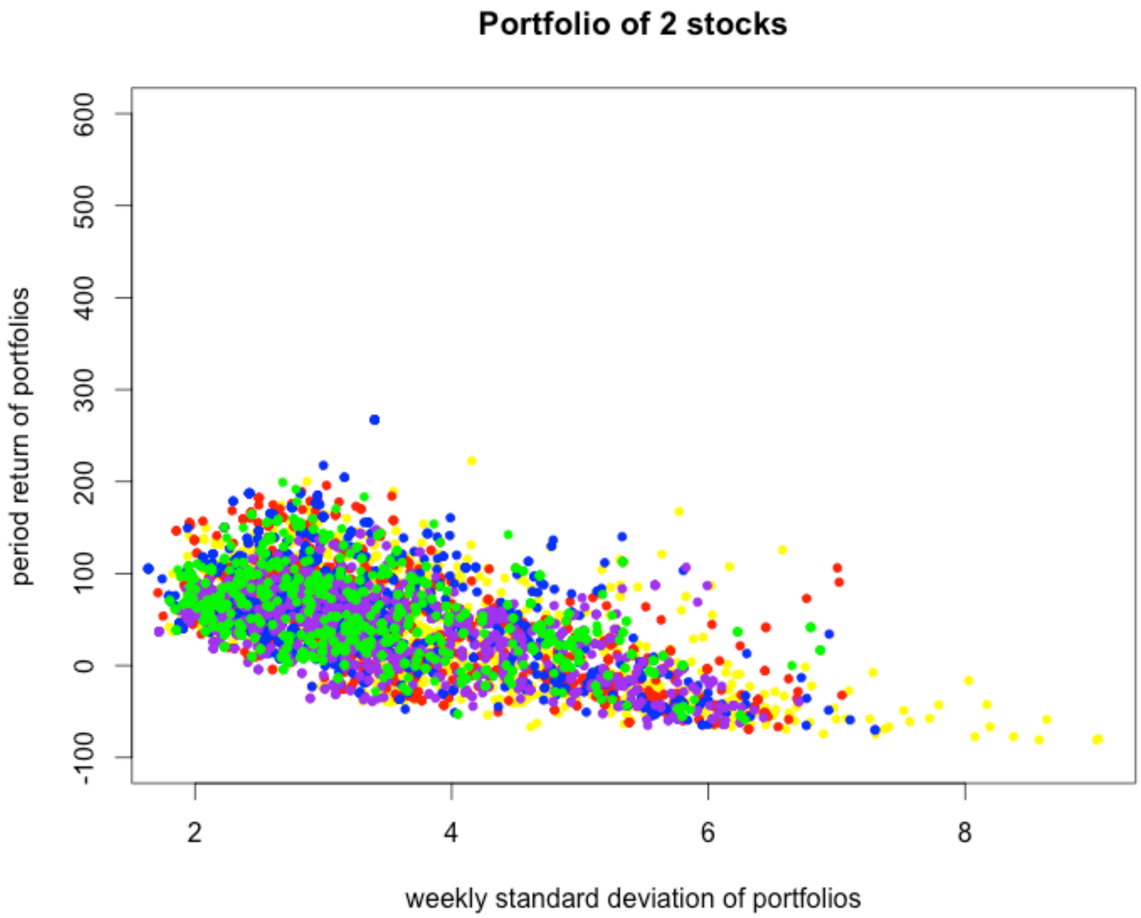


Figure 6.31 (e). The histogram and density curve of the simulated portfolios' percentage returns for period 5: 11 March 2009 to 5 October 2011. Each portfolio contains two stocks which were picked from different clusters that are bounded by the neighbor-Net split graph produced from the stocks' weekly returns in period 4: 17 October 2007 to 4 March 2009 but not industry groups.



- Stocks picked randomly.
- Stocks picked from different correlation clusters suggested by the neighbor-Net splits graph.
- Stocks picked from different industry groups.
- Stocks picked from different correlation and industry groups.
- Stocks picked from different correlation and non-industry groups.

Figure 6.32 Weekly standard deviation and return of each set of simulated 2-stock portfolios for period 5 using the five methods. The period returns and weekly standard deviations are expressed in percentage.

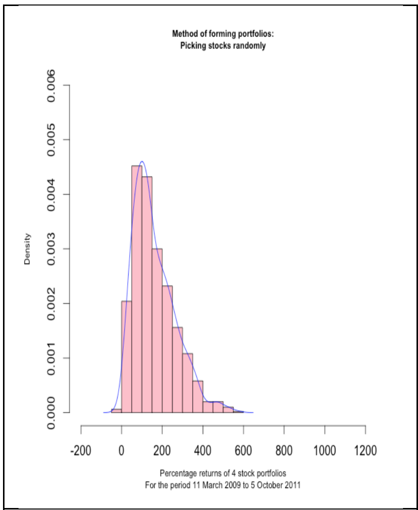


Figure 6.33 (a). The histogram and density curve of the simulated portfolios' percentage returns for period 5: 11 March 2009 to 5 October 2011. Each portfolio contains four stocks which were picked randomly.

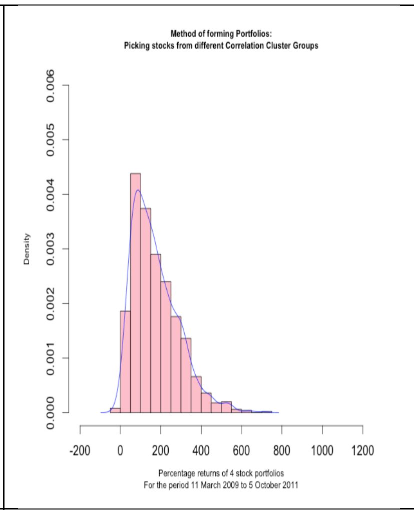


Figure 6.33 (b). The histogram and density curve of the simulated portfolios' percentage returns for period 5: 11 March 2009 to 5 October 2011. Each portfolio contains four stocks which were picked from different correlation clusters suggested by the neighbor-Net split graph produced from the stocks' weekly returns in period 4: 17 October 2007 to 4 March 2009.

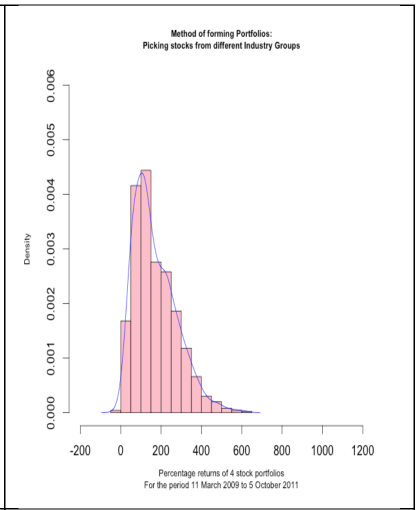


Figure 6.33 (c). The histogram and density curve of the simulated portfolios' percentage returns for period 5: 11 March 2009 to 5 October 2011. Each portfolio contains four stocks which were picked from different industry groups.

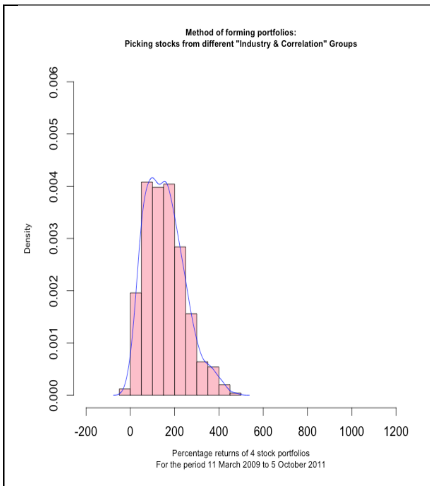


Figure 6.33 (d). The histogram and density curve of the simulated portfolios' percentage returns for period 5: 11 March 2009 to 5 October 2011. Each portfolio contains four stocks which were picked from different clusters that are bounded by both the neighbor-Net split graph produced from the stocks' weekly returns in period 4: 17 October 2007 to 4 March 2009.

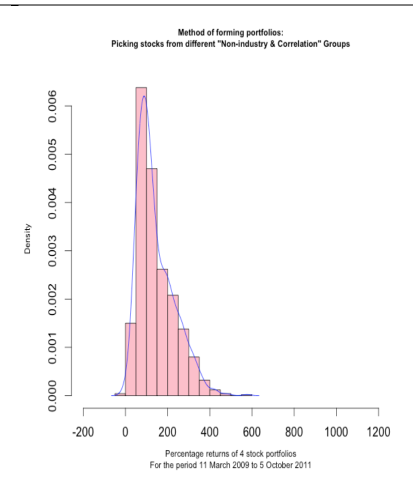
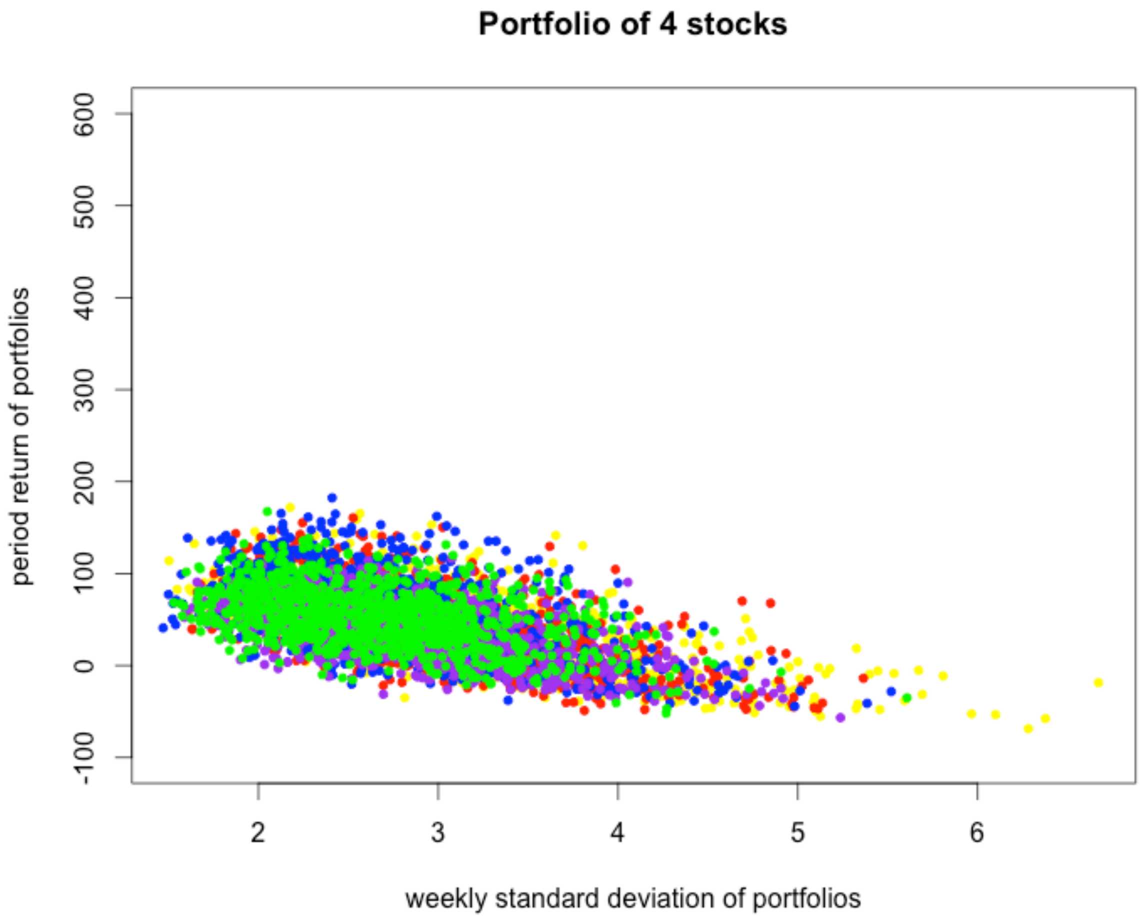


Figure 6.33 (e). The histogram and density curve of the simulated portfolios' percentage returns for period 5: 11 March 2009 to 5 October 2011. Each portfolio contains four stocks which were picked from different clusters that are bounded by the neighbor-Net split graph produced from the stocks' weekly returns in period 4: 17 October 2007 to 4 March 2009 but not industry groups.



- Stocks picked randomly.
- Stocks picked from different correlation clusters suggested by the neighbor-Net splits graph.
- Stocks picked from different industry groups.
- Stocks picked from different correlation and industry groups.
- Stocks picked from different correlation and non-industry groups.

Figure 6.34 Weekly standard deviation and return of each set of simulated 4-stock portfolios for period 5 using the five methods. The period returns and weekly standard deviations are expressed in percentage.

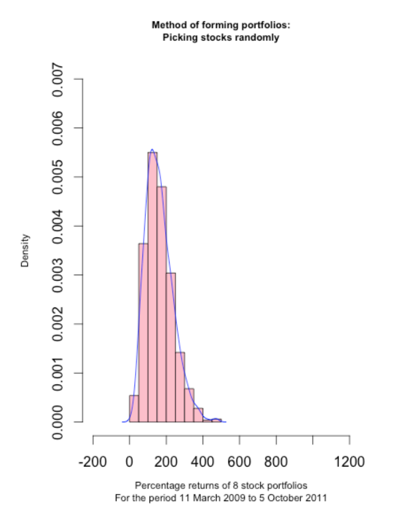


Figure 6.35 (a). The histogram and density curve of the simulated portfolios' percentage returns for period 5: 11 March 2009 to 5 October 2011. Each portfolio contains eight stocks which were picked randomly.

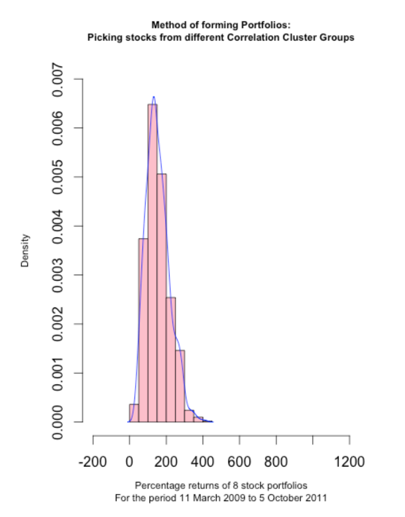


Figure 6.35 (b). The histogram and density curve of the simulated portfolios' percentage returns for period 5: 11 March 2009 to 5 October 2011. Each portfolio contains eight stocks which were picked from different correlation clusters suggested by the neighbor-Net split graph produced from the stocks' weekly returns in period 4: 17 October 2007 to 4 March 2009.

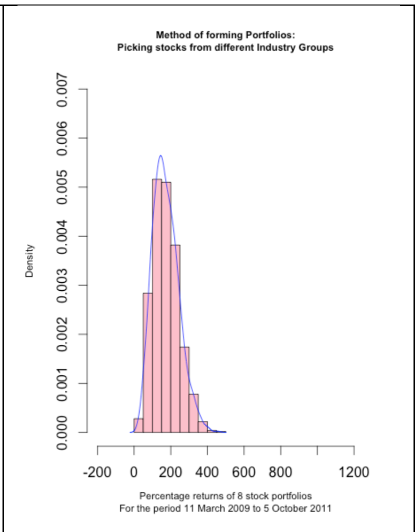


Figure 6.35 (c). The histogram and density curve of the simulated portfolios' percentage returns for period 5: 11 March 2009 to 5 October 2011. Each portfolio contains eight stocks which were picked from different industry groups.

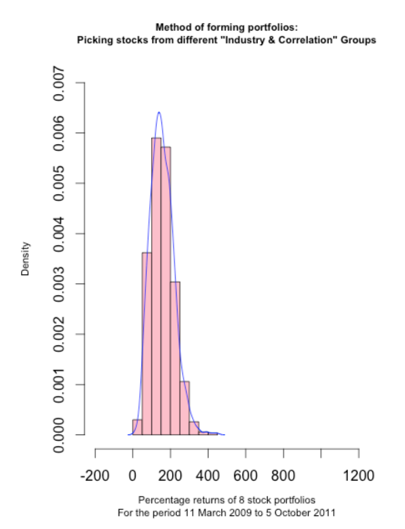


Figure 6.35 (d). The histogram and density curve of the simulated portfolios' percentage returns for period 5: 11 March 2009 to 5 October 2011. Each portfolio contains eight stocks which were picked from different clusters that are bounded by both the neighbor-Net split graph produced from the stocks' weekly returns in period 4: 17 October 2007 to 4 March 2009.

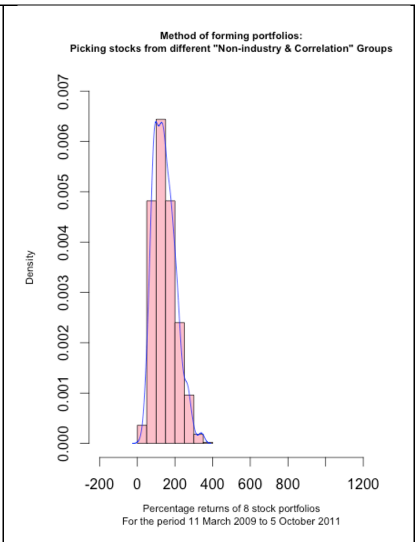
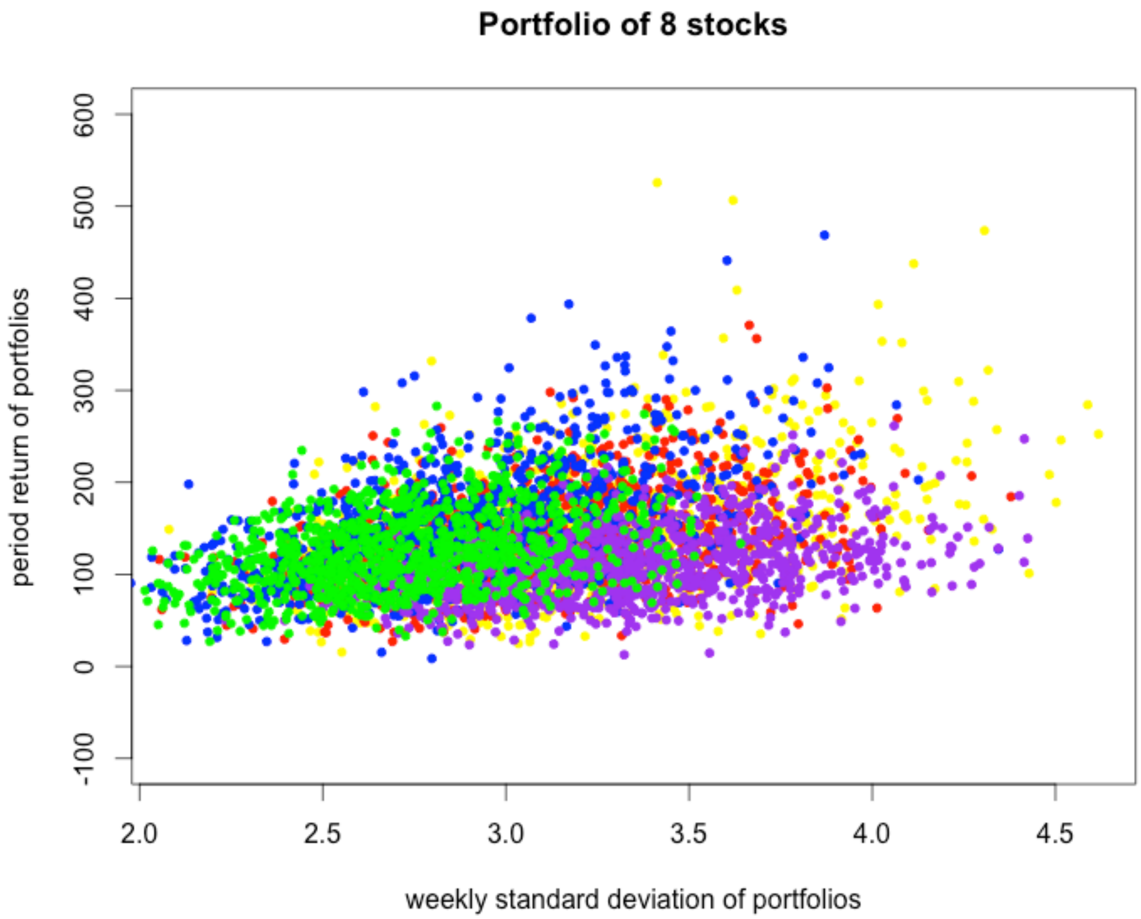


Figure 6.35 (e). The histogram and density curve of the simulated portfolios' percentage returns for period 5: 11 March 2009 to 5 October 2011. Each portfolio contains eight stocks which were picked from different clusters that are bounded by the neighbor-Net split graph produced from the stocks' weekly returns in period 4: 17 October 2007 to 4 March 2009 but not industry groups.



- Stocks picked randomly.
- Stocks picked from different correlation clusters suggested by the neighbor-Net splits graph.
- Stocks picked from different industry groups.
- Stocks picked from different correlation and industry groups.
- Stocks picked from different correlation and non-industry groups.

Figure 6.36 Weekly standard deviation and return of each set of simulated 8-stock portfolios for period 5 using the five methods. The period returns and weekly standard deviations are expressed in percentage.

Period 6

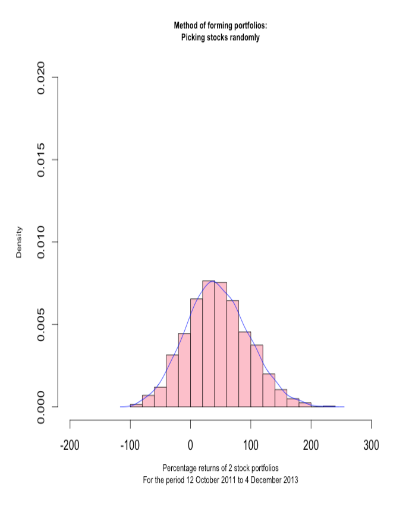


Figure 6.37 (a). The histogram and density curve of the simulated portfolios' percentage returns for period 6: 12 October 2011 to 4 December 2013. Each portfolio contains two stocks which were picked randomly.

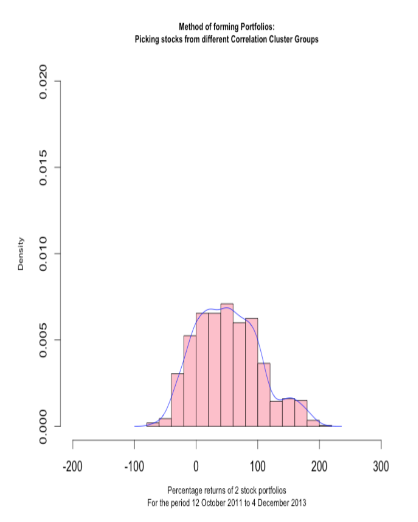


Figure 6.37 (b). The histogram and density curve of the simulated portfolios' percentage returns for period 6: 12 October 2011 to 4 December 2013. Each portfolio contains two stocks which were picked from different correlation clusters suggested by the neighbor-Net split graph produced from the stocks' weekly returns in period 5: 11 March 2009 to 5 October 2011.

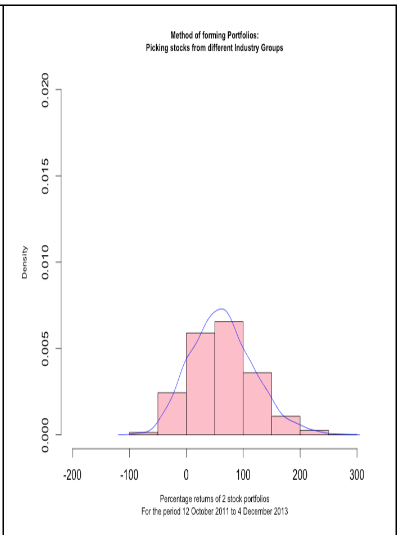


Figure 6.37 (c). The histogram and density curve of the simulated portfolios' percentage returns for period 6: 12 October 2011 to 4 December 2013. Each portfolio contains two stocks which were picked from different industry groups.

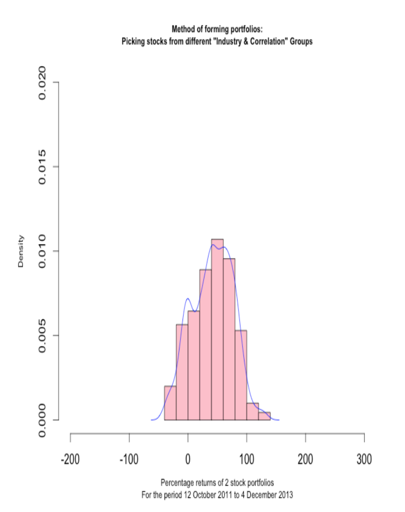


Figure 6.37 (d). The histogram and density curve of the simulated portfolios' percentage returns for period 6: 12 October 2011 to 4 December 2013. Each portfolio contains two stocks which were picked from different clusters that are bounded by both the neighbor-Net split graph produced from the stocks' weekly returns in period 5: 11 March 2009 to 5 October 2011.

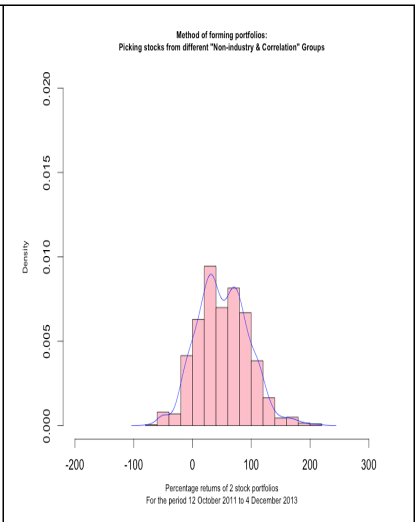
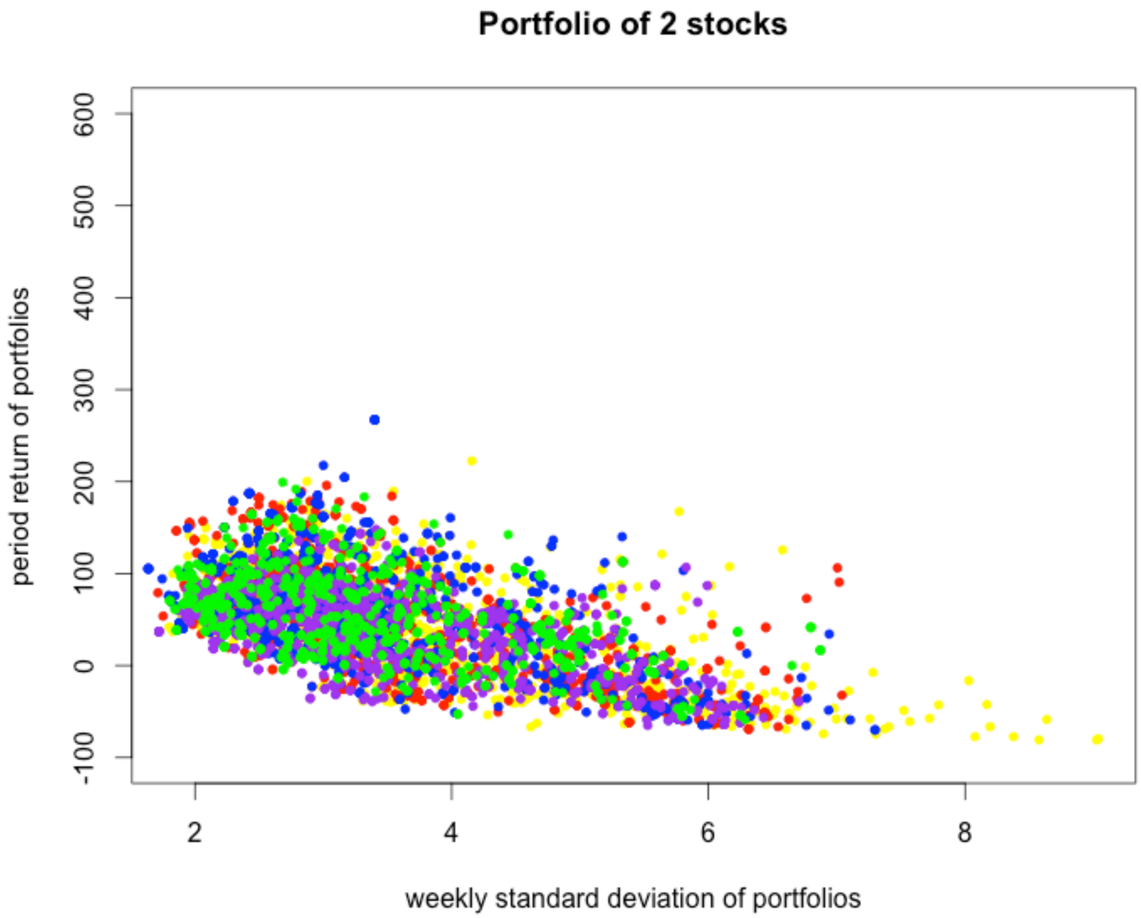


Figure 6.37 (e). The histogram and density curve of the simulated portfolios' percentage returns for period 6: 12 October 2011 to 4 December 2013. Each portfolio contains two stocks which were picked from different clusters that are bounded by the neighbor-Net split graph produced from the stocks' weekly returns in period 5: 11 March 2009 to 5 October 2011. but not industry groups.



- Stocks picked randomly.
- Stocks picked from different correlation clusters suggested by the neighbor-Net splits graph.
- Stocks picked from different industry groups.
- Stocks picked from different correlation and industry groups.
- Stocks picked from different correlation and non-industry groups.

Figure 6.38 Weekly standard deviation and return of each set of simulated 2-stock portfolios for period 6 using the five methods. The period returns and weekly standard deviations are expressed in percentage.

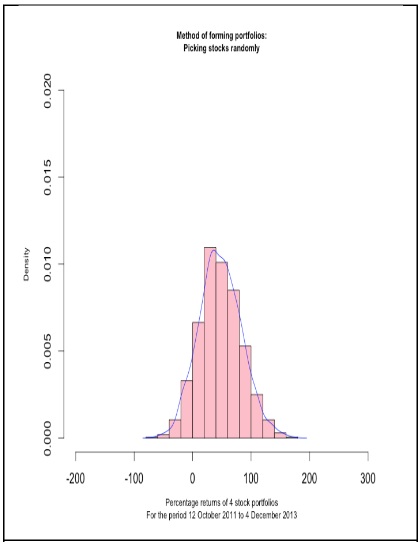


Figure 6.39 (a). The histogram and density curve of the simulated portfolios' percentage returns for period 6: 12 October 2011 to 4 December 2013. Each portfolio contains four stocks which were picked randomly.

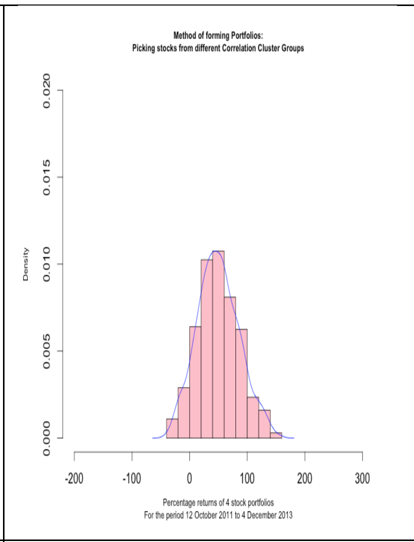


Figure 6.39 (b). The histogram and density curve of the simulated portfolios' percentage returns for period 6: 12 October 2011 to 4 December 2013. Each portfolio contains four stocks which were picked from different correlation clusters suggested by the neighbor-Net split graph produced from the stocks' weekly returns in period 5: 11 March 2009 to 5 October 2011.

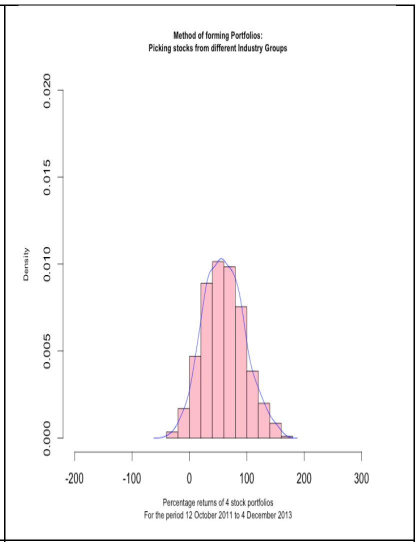


Figure 6.39 (c). The histogram and density curve of the simulated portfolios' percentage returns for period 6: 12 October 2011 to 4 December 2013. Each portfolio contains four stocks which were picked from different industry groups.

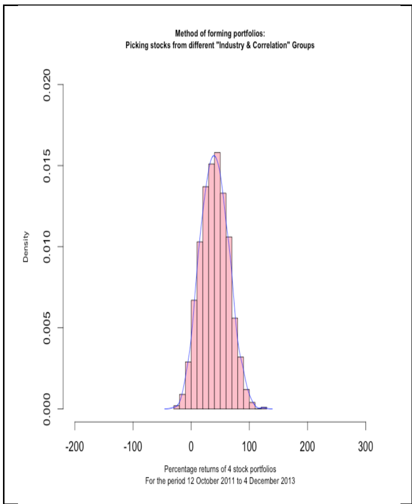


Figure 6.39 (d). The histogram and density curve of the simulated portfolios' percentage returns for period 6: 12 October 2011 to 4 December 2013. Each portfolio contains four stocks which were picked from different clusters that are bounded by both the neighbor-Net split graph produced from the stocks' weekly returns in period 5: 11 March 2009 to 5 October 2011.

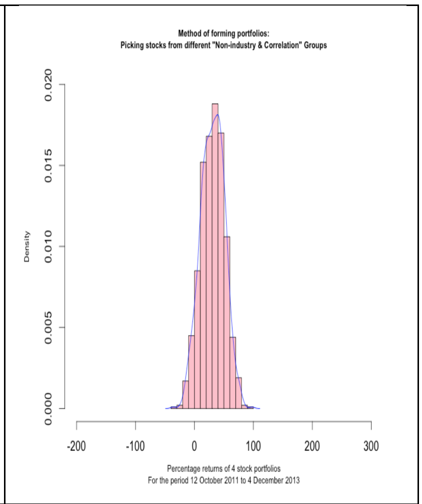
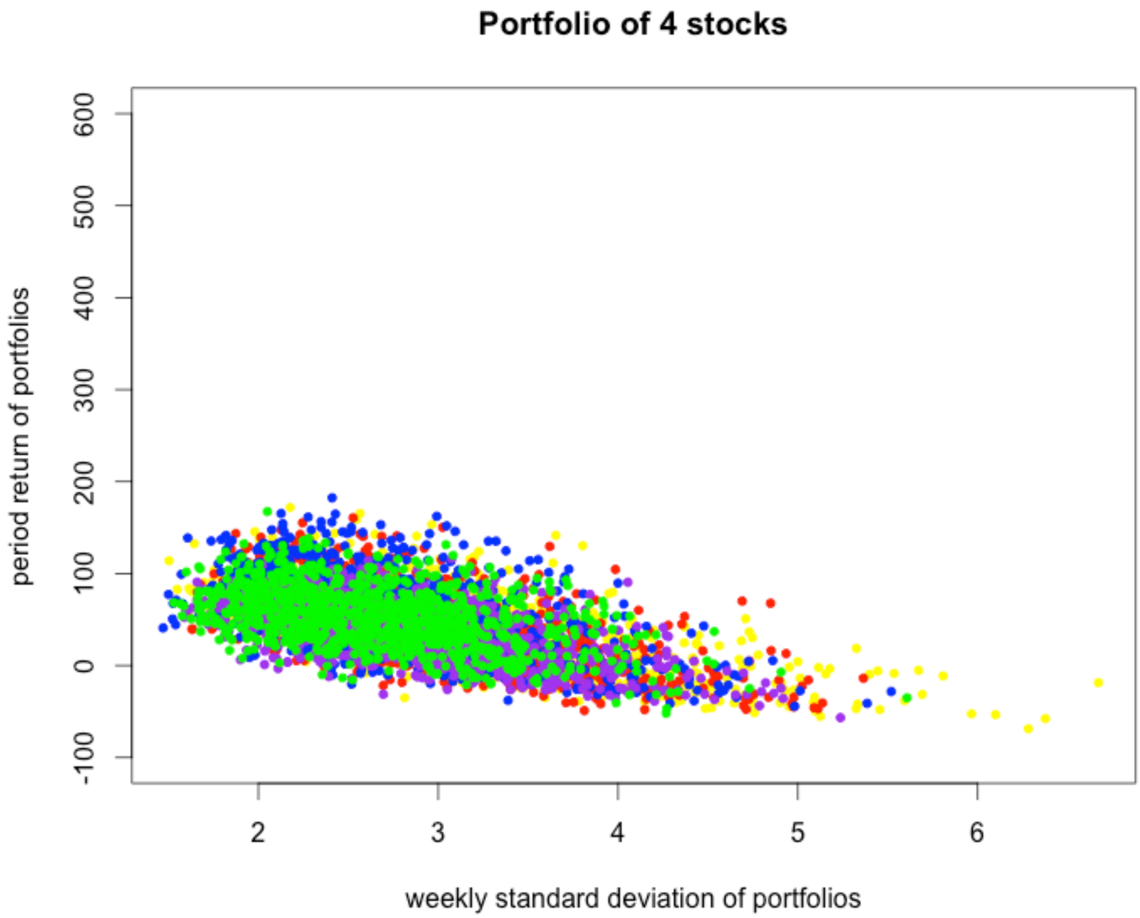


Figure 6.39 (e). The histogram and density curve of the simulated portfolios' percentage returns for period 6: 12 October 2011 to 4 December 2013. Each portfolio contains four stocks which were picked from different clusters that are bounded by the neighbor-Net split graph produced from the stocks' weekly returns in period 5: 11 March 2009 to 5 October 2011. but not industry groups.



- Stocks picked randomly.
- Stocks picked from different correlation clusters suggested by the neighbor-Net splits graph.
- Stocks picked from different industry groups.
- Stocks picked from different correlation and industry groups.
- Stocks picked from different correlation and non-industry groups.

Figure 6.40. Weekly standard deviation and return of each set of simulated 4-stock portfolios for period 6 using the five methods. The period returns and weekly standard deviations are expressed in percentage.

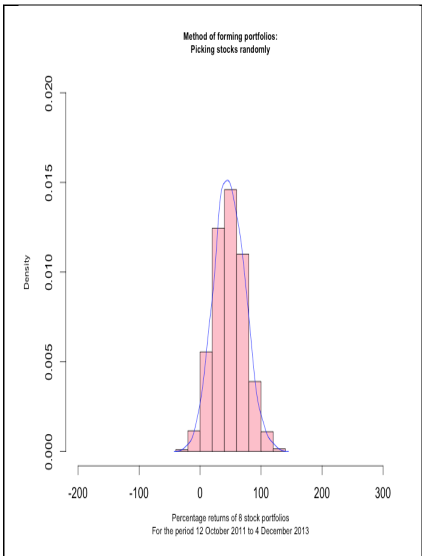


Figure 6.41 (a). The histogram and density curve of the simulated portfolios' percentage returns for period 6: 12 October 2011 to 4 December 2013. Each portfolio contains eight stocks which were picked randomly.

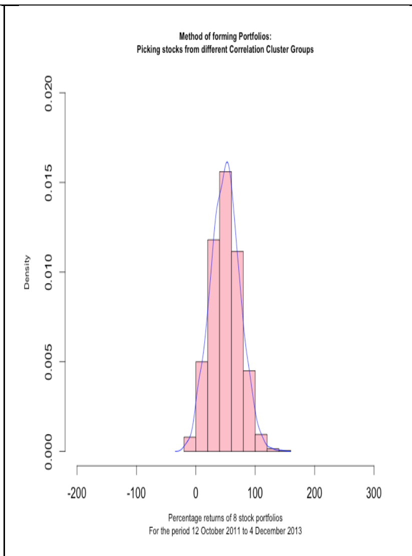


Figure 6.41 (b). The histogram and density curve of the simulated portfolios' percentage returns for period 6: 12 October 2011 to 4 December 2013. Each portfolio contains eight stocks which were picked from different correlation clusters suggested by the neighbor-Net split graph produced from the stocks' weekly returns in period 5: 11 March 2009 to 5 October 2011.

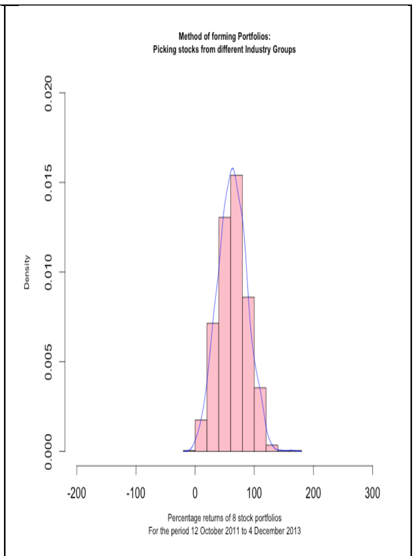


Figure 6.41 (c). The histogram and density curve of the simulated portfolios' percentage returns for period 6: 12 October 2011 to 4 December 2013. Each portfolio contains eight stocks which were picked from different industry groups.

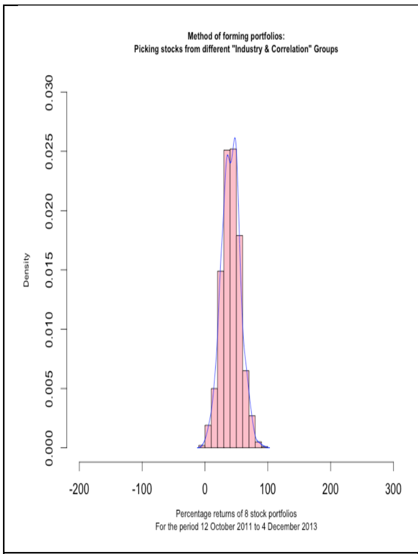


Figure 6.41 (d). The histogram and density curve of the simulated portfolios' percentage returns for period 6: 12 October 2011 to 4 December 2013. Each portfolio contains eight stocks which were picked from different clusters that are bounded by both the neighbor-Net split graph produced from the stocks' weekly returns in period 5: 11 March 2009 to 5 October 2011.

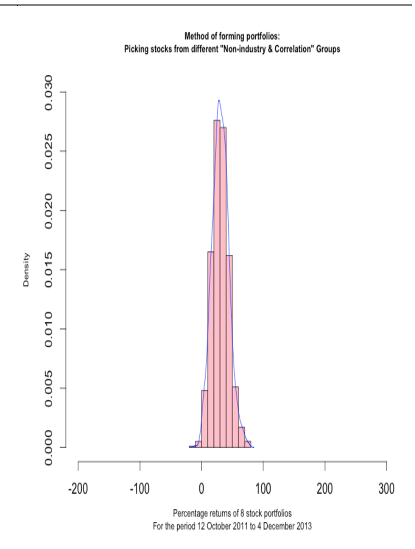
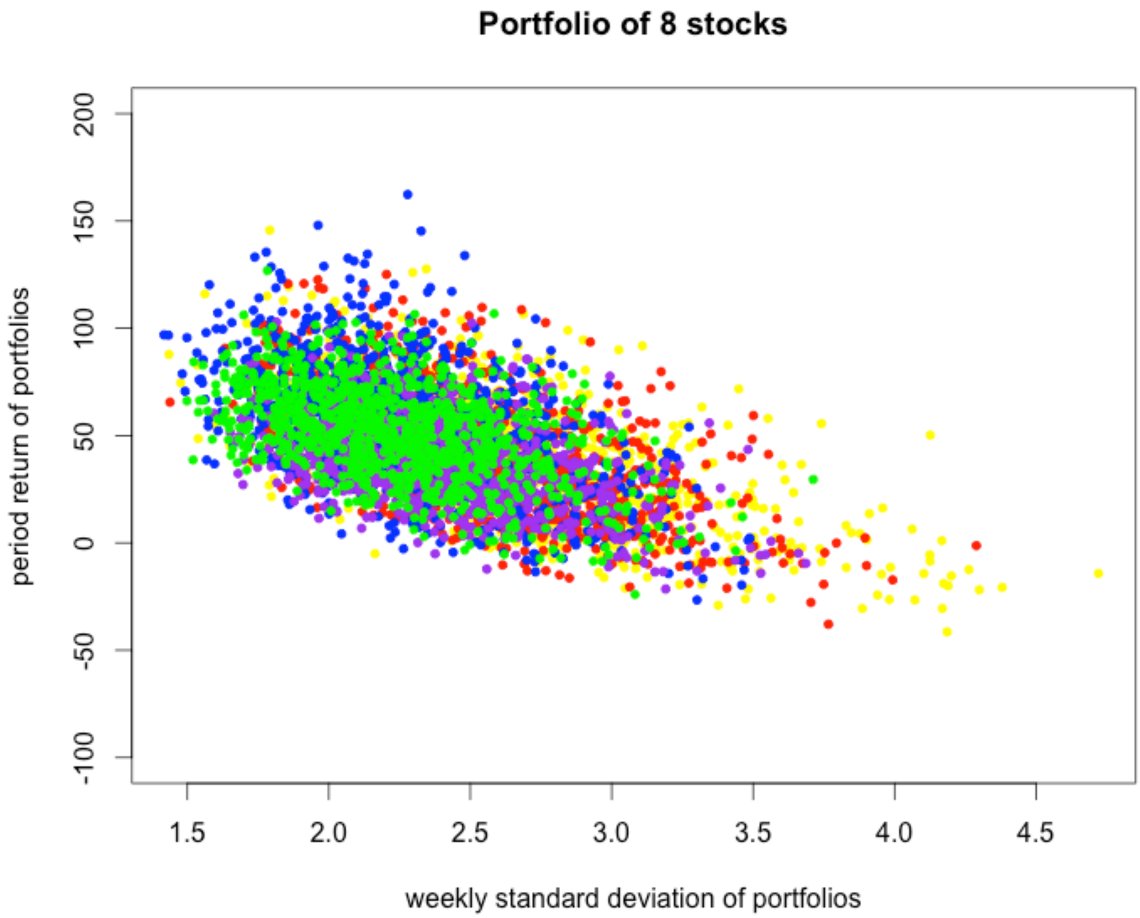


Figure 6.41 (e). The histogram and density curve of the simulated portfolios' percentage returns for period 6: 12 October 2011 to 4 December 2013. Each portfolio contains eight stocks which were picked from different clusters that are bounded by the neighbor-Net split graph produced from the stocks' weekly returns in period 5: 11 March 2009 to 5 October 2011. but not industry groups.



- Stocks picked randomly.
- Stocks picked from different correlation clusters suggested by the neighbor-Net splits graph.
- Stocks picked from different industry groups.
- Stocks picked from different correlation and industry groups.
- Stocks picked from different correlation and non-industry groups.

Figure 6.42 Weekly standard deviation and return of each set of simulated 8-stock portfolios for period 6 using the five methods. The period returns and weekly standard deviations are expressed in percentage.